Article

Organizational Research Methods
2020, Vol. 23(3) 399-431

DOI: 10.1177/1094428119840801
journals.sagepub.com/home/orm
SAGE

The Development and Construct Validity of a Team Processes Survey Measure

John E. Mathieu[1], Margaret M. Luciano[2], Lauren D'Innocenzo[3], Elizabeth A. Klock[1], and Jeffery A. LePine[2]

Abstract

Marks, Mathieu, and Zaccaro (2001) advanced a theory and framework of team processes that has garnered much attention and guided numerous investigations. They proposed that 10 first-order constructs (e.g., strategy formulation, coordination, conflict management) would map to three second-order constructs (i.e., transition, action, and interpersonal). Despite the popularity of this framework, we are unaware of any validated multiitem measures of the team processes they identified. Accordingly, we develop and demonstrate content and construct validity of 50-, 30-, and 10-item versions of a survey measure of team processes. Using data from over 700 teams, we test Marks et al.'s higher-order model and find results that are largely consistent with both their 10 first-order dimensions and the 10:3 second-order framework. Using samples of global virtual knowledge teams and health care employees, we provide evidence of the discriminant validity of our team process measure versus a measure of team empowerment. We provide recommendations for the use of these measures in future research and practice and encourage their use as part of a portfolio of measures of team processes.

Keywords

team process, scale development, construct validity

The Development and Construct Validity of a Team Processes Survey Measure

The past quarter-century has witnessed a surge in the use of teams in organizations along with a tremendous increase in research focused on their effectiveness (Mathieu, Hollenbeck, van

[1]University of Connecticut, Storrs, CT, USA
[2]Arizona State University, Tempe, AZ, USA
[3]Drexel University, Philadelphia, PA, USA

Corresponding Author:
John E. Mathieu, University of Connecticut, 2100 Hillside Road, Storrs, CT 06269, USA.
Email: john.mathieu@uconn.edu

Knippenberg, & Ilgen, 2017). Work teams are “interdependent collections of individuals who share responsibility for specific outcomes for their organizations” (Sundstrom, De Meuse, & Futrell, 1990, p. 120). Most of the recent work team research has been guided by the classic input-process-output (IPO) model of team effectiveness (cf. Hackman & Morris, 1975; McGrath, 1964). Mathieu, Heffner, Goodwin, Salas, and Cannon-Bowers (2000) characterized inputs as member, team, and organizational characteristics that exist prior to team performance episodes. Performance episodes are distinguishable periods of time when teams are performing taskwork. Processes describe how team inputs are transformed into outcomes that are the by-products of teamwork that are valued by one or more constituencies. Since its inception, the IPO framework has served as a unifying framework for research on small groups and teams. For instance, Salas, Stagl, Burke, and Goodwin (2007) reviewed the extant literature and identified 138 models that investigated three or more variables as related to team performance or effectiveness. A core theme of those investigations was that members engage in some form(s) of processes to transform inputs to outcomes. Work in the decade since has seen only a proliferation of more models of team effectiveness featuring one or more types of team processes (Mathieu, Wolfson, & Park, 2018).

Marks, Mathieu, and Zaccaro (2001) advanced a framework of team processes that has garnered much attention and guided numerous empirical investigations. They proposed that 10 first-order constructs (e.g., strategy formulation, coordination, conflict management) would map (10:3) to three second-order constructs (i.e., transition, action, and interpersonal) and represent how those team processes evolve over time. Despite the popularity of this model, to date, we are unaware of any validated multiitem measures of the team processes that they identified. Although there are many existing measures of team processes, they were not designed to tap the Marks et al. framework. Beyond the differences in labeling of the process measures, the content of the items is not completely consistent with the Marks et al. constructs, and thus the measures have potential issues with deficiency and contamination. For example, Saavedra, Earley, and Van Dyne’s (1993) measure of group task strategy is similar to Marks et al.’s strategy formulation and planning; however, the Saavedra measure includes content that implies creativity/imagination in devising the strategy. Similarly, the items in Drach-Zahavy and Somech’s (2001) measure of group learning seem to tap monitoring progress toward goals; however, the emphasis of the monitoring is somewhat unclear. Finally, there are scales that seem to capture interpersonal processes, but these tend to focus on the affective-cognitive-behavioral states themselves rather than the management of them. For instance, Hyatt and Ruddy’s (1997) work group confidence measure is similar to Marks et al.’s motivation and confidence building, but it reflects members’ confidence rather than activities meant to bolster members’ confidence. In short, none of the existing measures of team processes were designed specifically to align with Marks et al.’s framework, and evaluation in that light renders their suitability and construct validity unknown.

Accordingly, we have three goals for this article. First, we describe the development of multiitem scales for the Marks et al. (2001) processes and provide evidence of their content validity and internal consistency. Second, we test the second-order nature of Marks et al.’s framework. Third, we demonstrate the discriminant validity of these measures of team processes from a well-established measure of a team emergent state (i.e., empowerment). We conclude with recommendations for future research and application of these and other measures of team processes.

Marks et al. (2001) Framework

Conceptualizing Team Processes

Marks et al. (2001) defined team process as:

> members' interdependent acts that convert inputs to outcomes through cognitive, verbal and behavioral activities directed toward organizing taskwork to achieve collective goals . . . . Centrally, team process involves members interacting with other members and their task. They are the means by which members work interdependently to utilize various resources such as expertise, equipment, money, to yield meaningful outcomes (e.g., product development, rate of work, team commitment, satisfaction). (p. 357)

Team processes, like many constructs in industrial/organizational and organizational behavior, such as procedural justice (e.g., Colquitt, 2001), core self-evaluations (e.g., Judge, Erez, Bono, & Thoresen, 2003), team empowerment (e.g., Kirkman & Rosen, 1999), Big 5 personality (e.g., Digmann, 1997), and job satisfaction (e.g., Gillet & Schwab, 1975), have been conceptualized at different levels of specificity or granularity (see Bagozzi & Edwards, 1998). Team processes may be conceptualized as an overall composite, or the tendency of teams to function well with respect to transforming inputs into outcomes, or more narrowly in terms of more specific processes, or modes of functioning that manifest in response to different opportunities or situations. For instance, Marks et al. (2001) submitted that teams exhibit different processes at different times as related to performance episodes. They argued that prior to and between performance episodes, teams execute *transition processes* where they review their previous efforts, interpret their environment, and prepare for future work. They considered *action processes* as behaviors that team members engage in during episodes when their primary work is accomplished. Finally, they argued that teams need to continually manage *interpersonal processes* over time. Marks et al. then detailed more specific processes that fell primarily within each of these domains and thereby advanced a higher-order framework. LePine, Piccolo, Jackson, Mathieu, and Saul (2008) reviewed and categorized the teams literature using the Marks et al. framework and reported meta-analytic results showing that team processes exhibit positive relationships with team performance, cohesion, and potency as well as members' satisfaction.

Marks et al.'s (2001) framework has a regulatory focus theme and features performance episodes. Performance episodes represent meaningful periods of activities when work is accomplished and members take inventory and evaluate how well they are meeting their goals. Similar to goals, longer-term performance episodes (e.g., quarterly performance targets, patient care) may be built on shorter-term or subgoals (e.g., weekly goals, project completion, surgeries). Notably, the precise duration may vary, and tasks can be performed concurrently such that action processes for Task 1 may temporally co-occur with the transition processes for Task 2 (see Marks et al., 2001, Figure 1, p. 361). Although the precise duration may vary, the key element is that episodes focus attention on preparation and execution cycles.

Transition processes occur prior to or between performance episodes and have a dual focus whereby members reflect on and interpret previous accomplishments as well as prepare for future actions. Naturally during inception (e.g., a project team launch), initial transition processes have only a future orientation, but thereafter, they occur between performance episodes. Marks et al. (2001) identified three primary transition processes: (a) *mission analysis*—the identification and evaluation of team tasks, challenges, environmental conditions, and resources available for performing the team's work; (b) *goal specification*—activities centered on the identification and prioritization of team goals; and (c) s*trategy formulation and planning*—developing courses of actions and

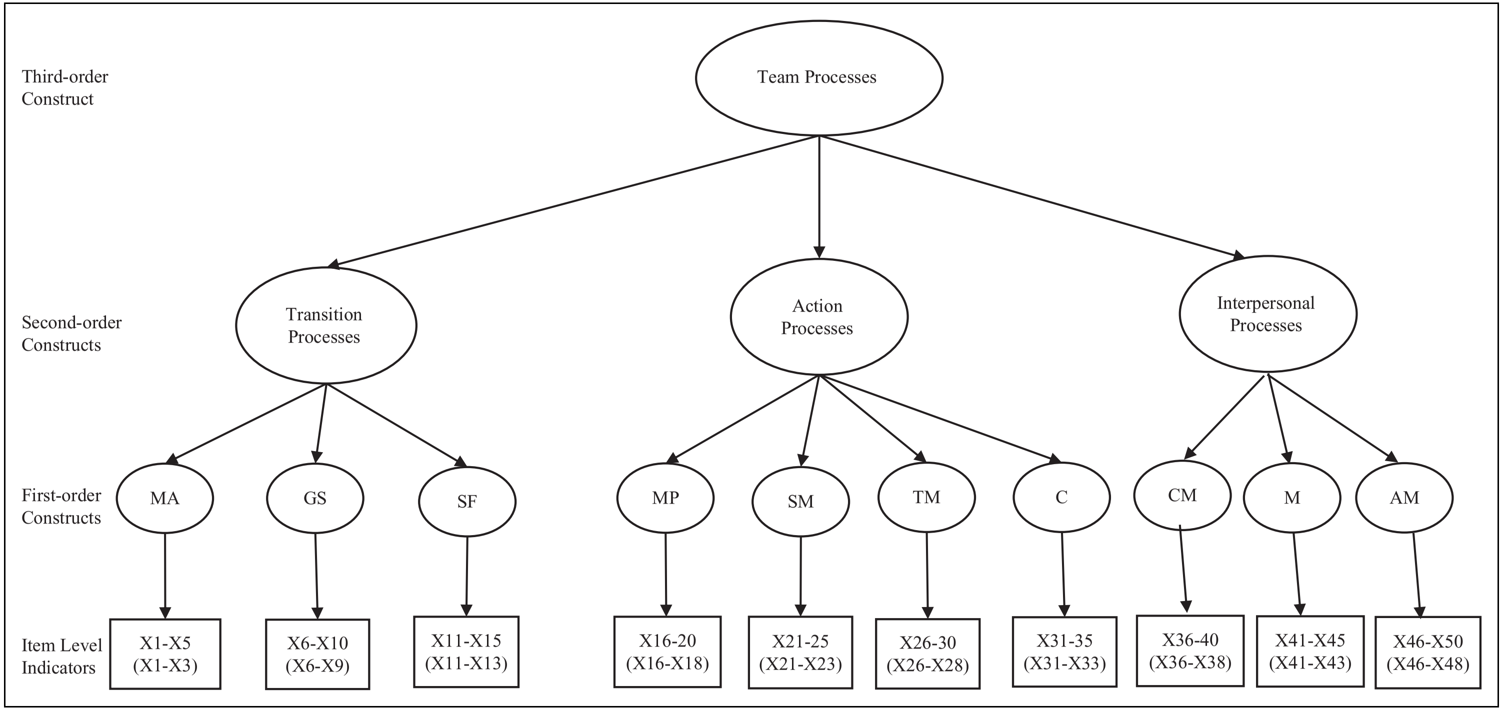

Figure 1. Marks, Mathieu, and Zaccaro's (2001) higher-order team process framework.
Note: MA = mission analysis; GS = goal specification; SF = strategy formulation and planning; MP = monitoring progress toward goals; SM = systems monitoring; TM = team monitoring and backup responses; C = coordination; CM = conflict management; M = motivation and confidence building; AM = affect management.

contingency plans as well as making adjustments to plans in light of changes or expected changes in the team's environment.

Action processes describe the behaviors that members engage in while working toward goal accomplishment. Marks et al. (2001) identified four primary types of action processes: (a) *monitoring progress toward goals*—members paying attention to, interpreting, and communicating information necessary for the team to gauge its progress toward its goals; (b) s*ystems monitoring*—activities such as tracking team resources (e.g., money) and factors in the team environment (e.g., inventories) to ensure that the team has what it needs to accomplish its goals and objectives; (c) *team monitoring and backup*—members assisting others in the performance of their tasks (by providing feedback or coaching or assisting with the task itself); and (d) *coordination*—the process of synchronizing or aligning the members' actions.

The third domain of Marks et al.'s (2001) framework deals with *interpersonal processes* that focus on the personal relationships between members. Whereas transition and action processes cyclically follow one another over time, Marks and colleagues argued that managing the interpersonal dynamics among members is an ongoing activity over time. They identified three primary types of interpersonal processes: (a) *conflict management*—the manner in which team members proactively and reactively deal with conflict; (b) *motivating and confidence building*—activities that develop and maintain members' motivation and confidence while working toward team goals; and (c) *affect management*—activities that foster emotional balance, togetherness, and effective coping with stressful demands and frustration.

Although the Marks et al. (2001) theory suggests that teams generally execute different processes at different times, it is also the case that such processes are ongoing and may occur at any given time. In other words, while there is a natural rhythm of transition processes setting the stage for action processes, which together with outcomes trigger future transition processes, the ebbs and flows of teamwork may require actions that reflect simultaneous processes. For instance, Marks et al. submitted that lacking suitable predetermined performance strategies, team members may need to improvise during action processes and undertake reactive strategy adjustments to accomplish their goals. Moreover, teams are often pursuing multiple goals simultaneously such that they must engage in transition and action processes simultaneously to orchestrate their various obligations. All this suggests that assessing their multifaceted team process at any given time may yield important information and doing so repeatedly can enable the modeling of their dynamic team processes. Nonetheless, this requires validated measures of team processes, which is the primary objective of our work here.

Measuring Team Processes

The Marks et al. (2001) framework has proven popular, as evidenced by it having been cited over 2,600 times in Google Scholar as of January 2018. We also traced the article citations using Scopus, which limits its search to academic journals, and it yielded a total of 928 citations as of the same period. For this project, we chose to conduct a systematic review of the team's literature using guidance provided in an INGRoup (Interdisciplinary Network of Group Research) publication (Baumann, 2013), where the focus was on identifying specific outlets for group research. Using this as our base, we took our initial list and limited the Scopus citations to those that appeared in the following 14 journals: *Academy of Management Journal*, *Academy of Management Review*, *Administrative Science Quarterly*, *Group Dynamics*, *Journal of Applied Psychology*, *Journal of Management*, *Journal of Organizational Behavior*, *Leadership Quarterly*, *Organization Science*, *Organizational Behavior and Human Decision Processes*, *Personnel Psychology*, *Leadership Quarterly*, *Small Group Research*, and *Strategic Management Journal*, which yielded 297 articles. We

Table 1. Studies Appearing in Team Outlets That Have Indexed One or More Mark, Mathieu, and Zaccaro's (2001) Dimensions.

	Transition Processes				Action Processes					Interpersonal Processes			
	Mai	GS[j]	SFP[k]	Overall	MP[l]	SM[m]	TM[n]	C[o]	Overall	CM[p]	MCB[q]	AM[r]	Overall
Behfar, Peterson, Mannix, and Trochim (2008)[g]				X						X			
Chen, Kanfer, DeShon, Mathieu, and Kozlowski (2009)[e]									X				
Courtright, McCormick, Mistry, and Wang (2017)[e]		X											
DeChurch and Haas (2008)[e]			X					X					
De Jong and Elfring (2010)[e]							X						
Eddy, Tannenbaum, and Mathieu (2013)[e]				X[b]					X[b]				X[b]
Firth, Hollenbeck, Miles, Ilgen, and Barnes (2015)[g]								X					
Fisher, Bell, Dierdorff, and Belohlav (2012)[e]								X					
Fisher (2014)[ce]			X	X				X					X
Kamphuis, Gaillard, and Vogelaar (2011)[g]								X					
Killumets, D'Innocenzo, Maynard, and Mathieu (2015)[e]										X	X	X	X[a]
Li and Liao (2014)[e]								X					
Li, Zhao, Walter, Zhang, and Yu (2015)[e]							X						
Lorinkova, Pearsall, and Sims (2013)[g]								X					
Marks, Mathieu, DeChurch, Panzer, and Alonso (2005)[dhf]	X	X	X	X[a]	X	X	X	X	X[a]				
Maruping, Venkatesh, Thatcher, and Patel (2015)[e]				X[b]					X[b]				X[b]
Mathieu and Schulze (2006)[df]	X[a]	X[a]	X[a]	X[b]						X[a]	X[a]	X[a]	X[b]
Mathieu and Taylor (2007)[e]				X[b]					X[b]				X[b]
Mathieu, Heffner, Goodwin, Cannon-Bowers, and Salas (2005)[f,h]	X[b]												
Mathieu, Maynard, Taylor, Gilson, and Ruddy (2007)[e]				X[b]					X[b]				X[b]

(continued)

Table 1. (continued)

	Transition Processes				Action Processes					Interpersonal Processes			
	Mai	GS[j]	SFP[k]	Overall	MP[l]	SM[m]	TM[n]	C[o]	Overall	CM[p]	MCB[q]	AM[r]	Overall
Mathieu, Gilson, and Ruddy (2006)[e]				X[b]					X[b]				X[b]
Pearsall and Venkataramni (2015)[e]			X										
Rapp and Mathieu (2007)[d,f]	X[a]	X[a]	X[a]	X[b]	X[a]	X[a]	X[a]	X[a]	X[b]	X[a]	X[a]	X[a]	X[b]
Rapp, Bachrach, Rapp, and Mullins (2014)[e]					X								
Rapp, Gilson, Mathieu, and Ruddy (2016)[e]				X[b]					X[b]				X[b]
Sonnentag and Volmer (2009)[d,e]		X											
Sui, Wang, Kirkman, and Li (2016)[e]								X					
Summers, Humphrey, and Ferris (2012)[e]								X					
Tekleab, Quigley, and Tesluk (2009)[e]										X			

Note: This table represents articles that have cited Marks, Mathieu, and Zaccaro (2001) and attempted to measure at least 1 of the 10 dimensions in the team process framework.
[a]Overall average of lower dimensions. [b]Higher-order measure of team process. [c]All measures not administered at once. [d]Multiple administrations of the measure. [e]Measured by use of a scale. [f]Measured using Behaviorally Anchored Rating scales (BARS). [g]Measured using unobtrusive methods. [h]Measured using raters. [i]Mission analysis. [j]Goal specification. [k]Strategy formulation and planning. [l]Monitoring progress toward goals. [m]Systems monitoring. [n]Team monitoring and backup response. [o]Coordination. [p]Conflict management. [q]Motivation and confidence building. [r]Affect management.

reviewed each of those articles and identified 29 studies that indexed one or more of Marks et al.'s dimensions, as summarized in Table 1.

As shown in Table 1, to date, six studies have indexed transition, action, and interpersonal processes using items aligned at the second-order level (i.e., Eddy, Tannenbaum, & Mathieu, 2013; Maruping, Venkatesh, Thatcher, & Patel, 2015; Mathieu, Gilson, & Ruddy, 2006; Mathieu, Maynard, Taylor, Gilson, & Ruddy, 2007; Mathieu & Taylor, 2007; Rapp, Gilson, Mathieu, & Ruddy, 2016) or by averaging single-item Behaviorally Anchored Rating scales (BARS) for each of the first-order dimensions (i.e., Rapp & Mathieu, 2007). Other investigations have used multiitem scales to measure one—three first-order dimensions (e.g., Courtright, McCormick, Mistry, & Wang, 2017; Killumets, D'Innocenzo, Maynard, & Mathieu, 2015), count measures (i.e., Firth, Hollenbeck, Miles, Ilgen, & Barnes, 2015; Kamphuis, Gaillard, & Vogelaar, 2011; Lorinkova, Pearsall, & Sims, 2013), or qualitative methods (i.e., Behfar, Peterson, Mannix, & Trochim, 2008). Most studies assessed team processes using surveys completed by team members (23 studies) or observer ratings (i.e., Marks, Mathieu, DeChurch, Panzer, & Alonso, 2005). In sum, whereas numerous studies have indexed one or more of Marks et al.'s (2001) dimensions, to date, no one has measured the 10 first-order dimensions using multiitem scales.

LePine et al. (2008) noted that the Marks et al. (2001) framework implicitly advanced a higher-order structure of team processes whereby the 10 first-order dimensions mapped to three second-order dimensions and thereby to an overall measure of team process (i.e., third-order construct). Figure 1 depicts this higher-order framework. Using meta-analytic–derived correlations, they found that "the data were very consistent with the hierarchical structure of team processes that we anticipated" (LePine et al., 2008, p. 293). However, as a limitation of their investigation, they noted:

> The number of studies with relationships among different aspects of teamwork processes was fairly small . . . . That is, scholars clearly have not conducted enough research on the construct validity of narrow [first-order] group processes. We suggest that if scholars wish to study narrow aspects of group process, research aimed at supporting their theoretical and empirical distinctiveness should come first. (p. 296)

Although the meta-analysis conducted by LePine and colleagues (2008) offers insights concerning the factor structure of the team process dimensions, the use of meta-analytically derived correlation matrices that mix study measures and results from a variety of circumstances are not uniform across variable pairings (Landis, 2013). Their meta-analyses included a mix of effect sizes from indirect measures that were not originally developed to assess team process as defined by Marks et al. (2001). Therefore, any factor analysis based on those correlations may not align well with the conceptual domain outlined by Marks and colleagues. Moreover, the different correlations comprising the synthetic matrix are built on differing numbers of teams from different investigations and implicitly violate the multivariate normality assumption of using confirmatory factor analysis (Cheung & Chan, 2005). The consequences of such a violation are unknown and essentially intractable, suggesting that empirical validation of the higher-order framework using original data is warranted. In sum, a test of the hypothesized factor structure using consistent measures that are directly and explicitly intended to represent the conceptual domain is needed to draw stronger inferences regarding the functioning of teams and related team phenomena.

It is important to emphasize that different researchers may wish to assess team processes at different levels of specificity for different purposes (see Luciano, Mathieu, Park, & Tannenbaum, 2018). As shown in Table 1, some researchers (e.g., Tekleab, Quigley, & Tesluk, 2009) focused on 1 of the 10 specific dimensions that Marks et al. (2001) identified, whereas others focused on the three higher-order dimensions (e.g., Mathieu et al., 2006) or even an overall composite (e.g., Eddy et al., 2013). Bagozzi and Edwards (1998) submitted that "decisions regarding construct depth must often be made in conjunction with decisions as to the dimensionality of the construct under examination" (p. 46). They go on to stress that "one construct depth is not necessarily superior to another, nor is a unidimensional conceptualization of a construct necessarily better than a multidimensional one. Much depends on the nature of the research question under study and the researcher's purposes" (p. 47). Accordingly, we evaluate the validity of team process measurement at different levels of specificity ranging from multiple items intended to measure a specific subdimension of team processes to Marks et al.'s 10 subdimension's relations to three higher-order constructs and thereby to an overall team process dimension.

Item Generation and Content Validity

Herein we discuss the item generation and establish the content validity of a measure of team processes based on the Marks et al. (2001) framework before testing the higher-order framework validity and scale discriminant validity using original item-level data. Content validity is defined by Haynes, Richard, and Kubany (1995) as the extent to which "elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose" (p. 238). In establishing the content validity of these measures, we followed the general recommendations set out by McKenzie, Wood, Kotecki, Clark, and Brey (1999) and Rubio, Berg-Weger, Tebb, Lee, and Rauch (2003).

Item Generation

Together with Michelle Marks, the first author reviewed previous measures of team processes available in the literature. We then independently drafted survey items for each of the 10 Marks et al. (2001) dimensions. We closely attended to Marks et al.'s definitions of each of the first-order constructs and sought to generate items aligned with those themes. Generally speaking, we sought to write a general encompassing item for each dimension followed by items that tapped the subthemes evident in each category definition. We then met and integrated, revised, and cycled our items until we settled on a 50-item set that we believed evidenced high content and face validity (see Appendix A).

Content Validity

Subject matter experts sample. We employed a methodology outlined by Anderson and Gerbing (1991) for assessing substantive (i.e., content) validity of items. Specifically, we solicited input from 12 well-renowned teams scholars and their students and colleagues. We sent them a handout consisting of an overview of the Marks et al. (2001) framework and a one-page list of the definitions of the 10 first-order dimensions for use while classifying items using an online survey. We asked them to list their affiliations, disciplinary background, position, or academic level and self-assess their knowledge of the Marks et al. taxonomy as: (a) I've read the handout but otherwise am not familiar with it, (b) I've read the handout and seen it before and/or read some work that referenced it, or (c) I consider myself quite familiar with the taxonomy. The online survey then presented judges with our 50 items (randomized per respondent) and asked them to categorize each item into one of the 10 first-order Marks et al. dimensions.

Our contacts forwarded our request to others such that we received responses from 87 individuals from 27 institutions, who ranged from undergraduates to SIOP and AOM Fellows. We eliminated seven undergraduates and five practitioners who expressed little familiarity with the taxonomy. The resulting sample ($N = 75$) was 49% PhD students and 51% professors, and 56% reported their discipline as organizational behavior, 40% as industrial/organizational psychology, and 4% other (e.g., communications, health care). There were no significant differences in response patterns across any of these subgroupings.

Results. Appendix B (available in the online version of the journal) contains detailed results concerning the 50 items. Specifically, on average, 78% ($SD = 18\%$) of subject matter experts (SMEs) assigned items to their intended first-order dimension. Item percentages ranged from 33% to 100%. Specifically, for transition processes scale items: (a) mission analysis mean = 64%, range, 44%-80%; (b) goal specification mean = 74%, range, 53%-98%; and (c) strategy and formulation planning mean = 64%, range, 44%-94%. For action processes scale items: (a) monitoring progress toward goal mean = 92%, range, 67%-96%; (b) systems monitoring mean = 83%, range, 65%-93%; (c) team monitoring mean = 61%, range, 33%-93%; and (d) coordination mean = 82%, range, 66%-99%. And for interpersonal processes scale items: (a) conflict management mean = 74%, range, 51%-100%; (b) motivation and confidence building mean = 86%, range, 70%-99%; and (c) affect management mean = 83%, range, 55%-96%. Considered in terms of the three second-order dimensions, items were accurately classified 53% to 100% of the time (mean = 92%, $SD = 11\%$). Notably, the first-order and second-order accuracy rates are all significantly ($p < .01$) greater than random baselines of 10% or 33%, respectively, except for the team monitoring Item No. 26, "Develop standards for acceptable team member performance," which was more often classified as belonging to goal specification.

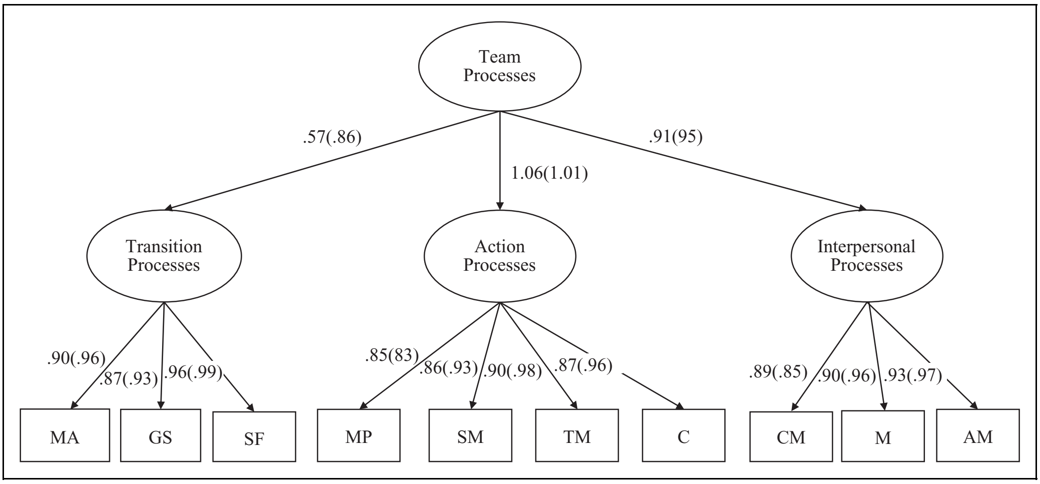

Figure 2. Fifty- and 30-item higher-order confirmatory factor analysis results.
Note: All parameter estimates are significant ($p < .001$). Fifty-item results outside parentheses, 30-item results within parentheses. Item loadings not shown.

Anderson and Gerbing (1991) describe the use of a *coefficient of substantive validity* (C_{SV}), which ranges from –1 to 1, with larger values indicating greater item substantive validity. More specifically, using the SMEs' item categorizations, items have a higher C_{SV} value to the extent that they are classified consistently as belonging to their intended construct as compared against their next most commonly used category. This provides a more stringent test than does a comparison against a random baseline. As shown in Appendix B (available in the online journal), besides the problematic item noted previously, five others were nonsignificant using this test. From mission analysis, No. 2, "Identify the key challenges that we expect to face," and No. 3, "Determine the resources that we need to be successful," were frequently classified as belonging to strategy formulation and planning and monitoring progress toward goals, respectively. "Periodically re-evaluate the quality of our working plans" (No. 14) from strategy formulation was more often classified as monitoring progress toward goals, while No. 38, "Maintaining group harmony," from conflict management was often assigned to affect management, and No. 46, "Share a sense of togetherness and cohesion," from affect management was often seen as indicative of motivation and confidence building. Interestingly, of the six nonsignificant items, three were attributed to other first-order dimensions within the same second-order dimension. The other three pertained to naturally aligned phenomena, such as determining and monitoring resources, monitoring goal progress and reevaluating plans, and monitoring members' behaviors against established standards. Marks et al. (2001) noted the inherent relationship between transition and action processes and actually foreshadowed such overlap and linkages (see their Figure 2, p. 364).

Psychometric Properties and Higher-Order Factor Structure

The content validity results chronicled previously are mainly supportive of the intended item mapping. SMEs consistently categorized most of the items according to their intended constructs. Ultimately, however, how team members respond to the items and the resulting psychometric properties and factor structure of the measures are critical information regarding their construct validity. Accordingly, in the following, we utilized data from numerous samples to evaluate the construct validity of different length versions of the measure.

Table 2. Overview of Samples in the Two 50- and Two 30-Item Analysis Subgroupings.

Team Types	Team Sizes	*N* Teams	*N* Members
50-Item Samples			
Field Investigations			
1. Baltic Region grocery chain, store operations	2-36 (*M* = 6.68, *SD* = 6.04)	82	548
2. Northeast USA grocery chain, cross-functional parallel teams	3-10 (*M* = 6.63, *SD* = 1.96)	46	305
Laboratory Investigations			
1. Southeast USA university, student classroom projects	2-5 (*M* = 3.33, *SD* = .96)	256	852
2. Singapore University, laboratory construction task	2-4 (*M* = 2.91, *SD* = .34)	123	356
30-Item Samples			
Field Investigations			
1. Manufacturing, global virtual knowledge teams	2-24 (*M* = 6.91, *SD* = 5.72)	32	221
2. USA aviation, production teams	6-18 (*M* = 11.05, *SD* = 3.79)	19	210
3. USA based, global business solutions teams	2-24 (*M* = 4.61, *SD* = 4.10)	44	203
Laboratory Investigations			
1. Southeast USA university, team SimCity simulation	3-member teams	26	78
2. Southeast USA university, simulated airport operations	7-12 (*M* = 9.86, *SD* = 1.25)	66	651
3. Southeast USA university, space flight simulation	3-member teams	20	60

Samples

The 50-item set of measures has been shared with numerous researchers for use in their empirical investigations in exchange for using their data to test the psychometric properties of the scales and their factor structure. Several investigators were not able to administer the entire set of items, so we identified a subset of 30 that we judged would function well as a short form. For that use, we selected items that (a) were more general and representative of the respective first-order constructs, (b) had demonstrated better psychometric properties in our initial applications, and (c) were preferred by practitioners and early users. In other words, the short form items were identified a priori by the authors long before we had amassed most of the samples used for the empirical results that follow. The short form items are the first three listed under each of the first-order dimensions in Appendix A. All items used a common lead-in stem asking, "To what extent does our team actively work to . . . ," and respondents used a 5-item response scale where 1 = *not at all*, 2 = *very little*, 3 = *to some extent*, 4 = *to a great extent*, and 5 = *to a very great extent*. In essence, respondents were reporting the extent to which their team engaged in effective team processes over some period in time. The period in question ranged from a previous experimental trial to the past 6 months, as applicable per sample.

Useable anonymized data were received from 10 samples, as detailed in Table 2. The samples range from students performing classroom and laboratory team tasks, to aviation production teams, to global business solutions knowledge teams. For use in analyses, we only included teams for whom we had two or more surveys. We also separated the samples by setting and survey version for comparative and cross-validation purposes: field (N = 853 members of 128 teams) and laboratory (N = 1,210 members of 379 teams), who completed the 50-item version of the survey, and field (N = 634 members of 95 teams) and laboratory (N = 789 members of 112 teams), who completed the 30-item version of the survey.

Analysis Strategy

We used the four subsamples to test the psychometric properties of both the items and the first-order scales as well as the fit of Marks et al.'s (2001) second-order process framework. We calculated intraclass correlations (ICCs) for the items and first- and second-order dimensions. ICC1 is the reliability of an individual's rating of the group mean, which is also the proportion of total variance attributable to team membership. In contrast, ICC2 represents the reliability of the group average rating (Chen, Mathieu, & Bliese, 2004). Notably, Bliese, Maltarich, Hendricks, Hofmann, and Adler (2018) "propose that group-level measurement validation be augmented with information about each scale item's ability to differentiate groups" (p. 1), which we provide in Appendix B (available in the online version of the journal). LeBreton and Senter (2008) suggested that ICC1 values of .01 might be considered small, $\geq$.10 as medium, and $\geq$.25 as large.

Although ICCs provide information about the relative between-team variance inherent in individual members' ratings, they are a type of interrater reliability and are not equivalent to and should not be confused with measures of agreement or internal consistency (LeBreton & Senter, 2008). The primary determinant of whether evidence exists to support aggregating individuals' referent-shift style measures (as we have here) are agreement indices (Chen et al., 2004; James, 1982). To wit, James (1982) submitted that "the use of aggregates for this purpose is predicated on demonstrating perceptual agreement because agreement implies a shared assignment of psychological meaning and provides a composition theory for climate [or processes in this case] at different levels of analysis" (p. 228). James went on to demonstrate that one cannot unequivocally attribute low ICCs to a lack of agreement because they are driven in large part by between-group variance. That was the reasoning behind James, Demaree, and Wolf's (1984) development of the r_{wg} index. Accordingly, given that these scales are referent shift (i.e., wording aligned with team-level processes), we report James et al.'s r_{wg} agreement index to justify aggregating individual members' responses to the team level. Notably, we applied this both to individual items and the first- and second-order dimensions versus the traditional null rectangular distribution. We adopted the rectangular distribution for these tests because we saw no basis for anticipating skewed distributions, nor were they prominent in the data. Traditionally, median r_{wg} values >.70 for multiitem scales are considered sufficient agreement to justify aggregation. To our knowledge, thresholds for individual items' medians have not been established, so we used the .60 value as a general guide. We calculated the team-level internal consistencies using the average item response per team as the inputs. This strategy aligns the measurement reliability information with the level of analyses used for substantive tests (see Chen et al., 2004).

Because our analyses are performed at the team level of analysis, the four subsamples listed in Table 2 are too small to perform confirmatory factor analyses (CFA) at the item level. Therefore, we first fit a single-factor solution to each of the five-item averages, per first-order dimension, to test for unidimensionality and convergent validity. Following the single-factor item models, we computed a scale score per dimension by averaging the respective items and then used the 10 scales (i.e., substantive parcels) as indicators to test the three-factor second-order model (10:3). Next, we combined the 50-item field and laboratory subsamples to yield a sufficient sample size for analyzing the second-order (50:10:3) framework using items as depicted in Figure 1. We also include comparative models to provide nested model tests of the discriminant validity of different solutions. Finally, we provide parallel analyses using the 30-item field and laboratory subsamples and then the combined samples.

To gauge model fit for the CFA analyses, we report the standardized root mean square residual (SRMR) and the Comparative Fit Index (CFI). We also report chi-square values, which provide a statistical basis for comparing the relative fit of nested models. We adopted the following guidelines for model fit suggested by Mathieu and Taylor (2006): Models with CFI values <.90 and SRMR

Table 3. Subscale Psychometric Properties.

	Field 50-Item				Lab 50-Item				Field 30-Item				Lab 30-Item			
	α	r_{wg}	ICC1	ICC2	α	r_{wg}	ICC1	ICC2	α	r_{wg}	ICC1	ICC2	α	r_{wg}	ICC1	ICC2
Transition	**.98**	**.97**	**.40**	**.82**	**.94**	**.97**	**.14**	**.34**	**.96**	**.94**	**.20**	**.63**	**.95**	**.93**	**.13**	**.52**
Mission analysis	.96	.92	.32	.75	.90	.94	.14	.34	.95	.87	.21	.63	.89	.89	.10	.45
Goal specification	.97	.94	.43	.83	.91	.92	.11	.29	.88	.82	.12	.46	.92	.83	.12	.50
Strategy formulation	.95	.91	.30	.74	.87	.92	.13	.33	.93	.84	.23	.66	.84	.79	.10	.44
Action	**.94**	**.98**	**.03**	**.19**	**.94**	**.97**	**.15**	**.36**	**.96**	**.95**	**.18**	**.60**	**.92**	**.95**	**.09**	**.41**
Mon. Goal progress	.89	.90	.06	.29	.89	.91	.11	.28	.91	.79	.21	.64	.79	.79	.04	.21
Systems monitoring	.84	.92	.05	.24	.91	.92	.14	.33	.86	.79	.16	.55	.75	.84	.05	.26
Team monitoring	.84	.90	.02	.11	.82	.92	.13	.32	.85	.77	.11	.45	.79	.80	.07	.33
Coordination	.86	.94	.02	.11	.93	.94	.17	.40	.86	.84	.16	.56	.94	.89	.15	.55
Interpersonal	**.96**	**.97**	**.05**	**.26**	**.96**	**.97**	**.20**	**.45**	**.96**	**.93**	**.12**	**.47**	**.93**	**.95**	**.13**	**.52**
Conflict management	.90	.94	.04	.21	.90	.94	.15	.35	.90	.79	.10	.43	.73	.89	.08	.39
Motivating and confidence	.89	.91	.03	.16	.93	.93	.20	.45	.89	.83	.11	.45	.92	.89	.16	.57
Affect management	.93	.91	.06	.29	.94	.93	.19	.43	.93	.78	.14	.52	.87	.87	.10	.45
N (individuals) teams	(853) 128				(1,210) 379				(634) 95				(789) 112			

Note: Bold values represent indices for the second-order dimensions.

values >.10 are *deficient*, those with CFI ≥.90 to <.95 and SRMR >.08 to ≤.10 are *acceptable*, and those with CFI ≥.95 and SRMR ≤.08 are *excellent*. Notably, of the two fit indices, deference is given to the SRMR values when working with relatively small samples size to parameter ratios, such as the number of teams to estimates here (Hu & Bentler, 1999). Finally, it is also the case that models with more constraints necessarily exhibit worse fit indices than models with fewer constraints (e.g., Marks et al.'s [2001] 10:3 second-order model vs. simply a 10-dimensional first-order model). Yet models with greater constraints offer more parsimonious explanations of relationships and, all else equal, are preferable (James, 1982; Mulaik, 1998). Accordingly, we also report a Parsimony Comparative Fit Index (PCFI) that takes into consideration the number of parameters in a model and its fit (i.e., CFI) and rewards models with relatively fewer degrees of freedom. Although there are no steadfast guidelines for interpretation, Hair, Black, Babin, and Anderson (2014) suggested parsimony values between .60 and .80 can be considered as good and values higher than .80 as excellent (also see Mulaik, 1998).

50-Item Version

Item and scale properties. The item-level analyses using the 50-item measure are summarized in Appendix B (available in the online version of the journal), and the first- and second-order scale properties are summarized in Table 3. For the field subsample, all individual items evidenced

Table 4. Second-Order Factor Loadings for Different Versions, Subsamples, and Analyses.

	50-Item Version			30-Item Version		
	Field Subscale[a] (10:3)	Lab Subscale[a] (10:3)	Combination Second Order[b] (50:10:3)	Field Subscale[a] (10:3)	Lab Subscale[a] (10:3)	Combination Second Order[b] (30:10:3)
Mission analysis	.970	.834	.902	.934	.903	.963
Goal specification	.949	.793	.875	.880	.914	.934
Strategy formulation and planning	.936	.885	.960	.933	.935	.986
Monitoring progress toward goals	.629	.824	.854	.794	.757	.832
System monitoring	.816	.831	.863	.904	.787	.931
Team monitoring and backup	.853	.789	.902	.898	.869	.984
Coordination	.896	.826	.868	.965	.824	.964
Conflict management	.868	.820	.892	.862	.779	.853
Motivating and confidence building	.874	.867	.900	.937	.898	.957
Affect management	.920	.883	.928	.904	.877	.969

Note: Table values are standardized factor loadings on intended latent constructs. All loadings are $p < .001$.
[a]Loadings using subscales as indicators ($df = 32$). [b]Second-order loadings using items as first-order indicators ($df = 1,162$).

median r_{wg}s between .53 and .83 (mean = .71). Only four items had median r_{wg}s <.60, and each was >.50. The 50-item field samples exhibited single-item ICC1s from .00 to .40 (mean = .12) and associated ICC2s from .00 to .82 (mean = .34). We then fit single-factor CFAs to each of the five-item sets per first-order dimension. The factor loadings from these analyses are presented in Appendix C (available in the online version of the journal), and the model fit indices are presented in Appendix D (available in the online version of the journal). Notably, all single-factor models exhibited acceptable to excellent fit (i.e., CFIs >.94, SMRM <.05), and all factor loadings were significant and >.50. All five-item scales evidenced median r_{wg}s $\geq$.90, αs >.80, ICC1s from .02 to .43 (mean = .13), and associated ICC2s from .11 to .83 (mean = .37).

For the laboratory subsample, the 50 individual items evidenced median r_{wg}s between .60 and .83 (mean = .73), single-item ICC1s from .00 to .26 (mean = .16), and associated ICC2s from .00 to .43 (mean = .29). The single-factor CFAs fit to each of the five-item sets for the lab samples exhibited acceptable to excellent fit (i.e., CFIs >.91, SMRM <.05, see Appendix D, available in the online version of the journal), and all factor loadings were significant and >.50 (see Appendix C, available in the online version of the journal). As shown in Table 3, the five-item scales evidenced median r_{wg}s $\geq$.90, αs >.80, ICC1s from .11 to .20 (mean = .15), and associated ICC2s from .28 to .45 (mean = .35) in the laboratory subsample.

Testing the higher-order structure. Given that the five-item scales exhibited high member agreement, unidimensionality, and acceptable internal consistencies, we averaged them per first-order dimension and then fit the 10 scale scores to a three-factor model (10:3) to test the Marks et al. (2001) implicit framework. The resulting factor loadings from these analyses are presented in Table 4, and the corresponding model fit indices are shown in Table 5. The three-factor model evidenced an acceptable fit using both the field, $\chi^2(32) = 95.48, p < .01$, CFI = .949, SRMR = .057, PCFI = .675, and laboratory, $\chi^2(32) = 355.20, p < .001$, CFI = .905, SRMR = .043, PCFI = .644, subsamples. Moreover, the three-factor models were significantly better than single-factor (10:1) models (which exhibited deficient fit) in both cases: field, $\Delta\chi^2(3) = 483.38, p < .001$; $\chi^2(35) = 578.86, p < .001$;

Table 5. Confirmatory Factor Analysis Second-Order Model Fit Indices.

Samples and Models	*df*	χ^2	CFI	SRMR	PCFI	$\Delta\chi^2$
50-Item Sample Models						
Field 10:1	35	578.86***	.562	.188	.437	
Field 10:3	32	95.48**	.949	.057	.675	483.38***[a]
Lab 10:1	35	429.41***	.884	.046	.688	
Lab 10:3	32	355.20***	.905	.043	.644	74.21***[a]
Combined 50:1	1,175	12,194.92***	.568	.129	.545	
Combined 50:10	1,130	4,137.99***	.882	.061	.814	8,056.93***[b]
Combined 50:10:1	1,165	5,258.48***	.839	.111	.798	
Combined 50:10:3	1,162	4,504.02	.869	.071	.824	754.46***[c]
30-Item Sample Models						
Field 10:1	35	254.07***	.816	.066	.635	
Field 10:3	32	130.81**	.917	.035	.652	123.26***[a]
Lab 10:1	35	132.05***	.910	.044	.708	
Lab 10:3	32	75.24***	.960	.033	.683	56.81***[a]
Combined 30:1	405	2,392.99***	.736	.072	.685	
Combined 30:10	360	1,238.33***	.883	.061	.731	1,154.66***[b]
Combined 30:10:1	395	1,582.86***	.842	.076	.765	
Combined 30:10:3	392	1,372.87***	.870	.065	.786	209.99***[c]
Using all 714 Teams						
Combined 30:1	405	8,033.63***	.634	.107	.590	
Combined 30:10	360	1,917.43***	.925	.051	.766	6,116.20***[b]
Combined 30:10:1	395	3,203.17***	.865	.093	.785	
Combined 30:10:3	392	2,427.84***	.905	.061	.816	775.33***[c]

Note: *df* = degrees of freedom; CFI = Comparative Fit Index; SRMR = standardized root mean squared residual; PCFI = Parsimony Comparative Fit Index.
[a]Versus 10:1 model (*df* = 3). [b]Versus 50(30):1 model (*df* = 45). [c]Versus 50(30):10:1 model (*df* = 3).
**$p < .01$. ***$p < .001$.

Table 6. Correlations Between Second-Order Transition, Action, and Interpersonal Dimensions for Different Models.

Samples	*N*	Transition-Action	Transition-Interpersonal	Action-Interpersonal
50-Item Samples				
Field (10:3)	128	.447	.174[a]	.918
Lab (10:3)	379	.975	.886	.926
Combined (50:10:3)	507	.602	.517	.956
30-Item Samples				
Field (10:3)	95	.827	.770	.957
Lab (10:3)	112	.955	.851	.971
Combined 30-Item Sample				
(30:10:3)	207	.866	.821	.957
Fully Combined Sample				
(30:10:3)	714	.691	.627	.966
LePine et al. (2008)				
Meta-analysis	941	.96	.77	.86

Note: Table values are correlations between estimated latent variables.
All values $p < .001$ except [a]*ns*.

CFI = .562, SRMR = .188, PCFI = .437; laboratory, $\Delta\chi^2(3) = 74.21, p < .001$; $\chi^2(35) = 429.41, p < .001$; CFI = .884, SRMR = .046, PCFI = .688; and evidenced good parsimony indices. The correlations between the estimated second-order factors from the 10:3 models are presented in Table 6 along with those from later models and the meta-analytic correlations found by LePine et al. (2008). Although empirically distinguishable,[1] the 50-item laboratory dimensions were highly correlated (>.85, $p < .001$), whereas the ones associating transition processes were noticeably lower for the 50-item field 10:3 model. We should note, however, that the two 50-item field subsamples both collected measures of transition processes at different times than the action and interpersonal processes. No doubt this at least partially accounts for the lower observed correlations. In sum, these findings are consistent with Marks et al.'s higher-order framework and parallel LePine et al.'s results based on a CFA of meta-analytic correlations among indirect measures of the 10 first-order dimensions.

Finally, combining the 50-item subsamples ($N = 507$ teams) yields sufficient degrees of freedom to test the second-order factor structure at the item level. Fitting the 50 items to a 10-factor first-order model (50:10) yielded a deficient overall fit, $\chi^2(1, 130) = 4{,}127.99, p < .001$, CFI = .882, SRMR = .061, PCFI = .814, on the basis of the CFI but not SRMR. The 10-factor model was significantly better than a single-factor (50:1) model, which illustrated a poor fit, $\Delta\chi^2(45) = 8{,}056.93, p < .001$; $\chi^2(1, 175) = 12{,}194.92, p < .001$; CFI = .568, SRMR = .129, PCFI = .545. The 50:10 model also evidenced an excellent parsimony index. Notably, the item loadings in this analysis were all >.55, $p < .001$, on their intended constructs. Fitting the 10 first-order constructs to the three second-order dimensions (50:10:3) yielded a significantly worse model fit than the 50:10 model, $\Delta\chi^2(32) = 376.03, p < .001$; $\chi^2(1, 162) = 4{,}504.02, p < .001$; CFI = .869, SRMR = .071, PCFI = .824, although the first-order dimensions all loaded significantly and >.85, $p < .001$, on their intended second-order constructs. Here again, although the CFI is below conventional levels of acceptability, the SRMR value suggests an acceptable fit, and the parsimony index was excellent. Worth noting, the 50:10:3 higher-order structure fit significantly better, $\Delta\chi^2(3) = 754.46, p < .001$, than a 50:10:1 structure, $\chi^2(1, 165) = 5{,}258.48, p < .001$; CFI = .839, SRMR = .111, PCFI = .798, providing further evidence in support of the discriminant validity of the three second-order dimensions. Nevertheless, the second-order variables were highly correlated (mean $r = .89, p < .001$; see Table 6). Finally, as shown in Figure 2, we fit a 50:10:3:1 model to parallel the results reported by LePine et al. (2008). As shown, all of the first- and second-order loadings were significant (>.80, $p < .001$), as were the second- and third-order loadings (>.55, $p < .001$).[2]

Summary of 50-item measure. The analyses summarized previously provided mostly but not complete support concerning the psychometric properties of the 10 five-item scales. At the item level of analysis, SMEs significantly categorized 88% of the items as belonging to their intended first-order dimension more so than the next most frequently chosen alternative. The six nonsignificant items were seen as belonging to other dimensions within the same second-order Marks et al. (2001) dimension or closely related dimensions (e.g., setting vs. monitoring progress toward goals). Interestingly, none of those six items evidenced poor factor loadings (all >.70, $p < .001$) in the CFA analyses in either the field or lab samples.

Bliese et al. (2018) noted that items with poor ICC1s can undermine the construct validity of measures of aggregate constructs. Interestingly, of the nine items that exhibited small ICCs <.10 (see LeBreton & Senter, 2008) with the laboratory samples, none had r_{wg}s <.60. Although 32 items evidenced small ICCs with the field sample, only 1 (No. 45) had an r_{wg} <.60. This suggests that the lower ICCs were more likely attributable to restricted ranges on those items than a lack of agreement among team members. None of the 50 items were consistently problematic when gauged on the combined bases of SMEs' categorizations, ICCs, and r_{wg}s, and all exhibited significant loadings on their intended constructs in the item-level CFA analyses.[3]

At the scale level of analyses, all of the five-item scales evidenced high agreement indices (r_{wg}s >.90), internal consistencies (αs >.90), and on average, medium ICCs (mean ICC1 = .14, ICC2 = .45). Although the ICC2 values might appear concerning, they are to be expected with teams that average around five members. Even large ICCs such as .25 would only yield an ICC2 of .63 in small groups ($N = 5$) research.

All five-item scales also exhibited acceptable CFA fit indices when fit to a single factor model, and the 10 first-order dimensions loaded significantly on their intended second-order dimensions when tested at the item (i.e., 50:10:3) or subscale (i.e., 10:3) levels. Although the first-order models (i.e., 50:10) fit significantly better than a second-order model (i.e., 50:10:3), the latter evidenced superior parsimony indices. These results support the use of the measures at the first- or second-order levels of analysis, depending on the purpose of any particular investigation.

Although the results summarized previously are largely consistent with expectations, there were some signs of problematic items. Moreover, many scholars and practitioners wish to use fewer items to assess these constructs. As noted previously, we had identified a priori a 30-item version of the scale that contained the items that we believed best reflected the core aspects of each of Marks et al.'s (2001) 10 first-order constructs. On the basis of the analyses of the 50-item samples, the 30 items included in the shorter form evidenced equal or better psychometric properties compared to the 20 items that were not included: Psa(1) = .76 versus .74; $C_{SV}(10)$ = .60 versus .60; ICC1 = .14 versus .13; ICC2 = .32 versus .30; r_{wg}= .76 versus .74; CFA loadings = .83 versus .78, for the mean included versus excluded items, respectively. Accordingly, in the following, we report the psychometric properties of the 30-item version using different subsamples.

30-Item Version

Item and scale properties. The item-level analyses are also summarized in Table 3 and Appendix C (available in the online version of the journal) and paralleled the findings for those respective items seen in the 50-item findings. Items loaded significantly ($p < .01$) and >.46 in single-factor CFAs.[4] At the first-order scale level of analysis, all three-item scales evidenced median r_{wg}s $\geq$.75, αs $\geq$.70, ICC1s from .10 to .23 (mean = .16), and associated ICC2s from .43 to .66 (mean = .54) with the field samples. For the laboratory samples, the three-item scales exhibited median r_{wg}s $\geq$75, αs $\geq$.70, ICC1s from .04 to .16 (mean = .10), and associated ICC2s from .21 to .57 (mean = .42). In sum, despite being 60% of the length of the 50-item version, the 30-item version exhibited comparable psychometric properties.

Testing the higher-order structure. We also tested Marks et al.'s (2001) higher-order structure using the samples that administered the 30-item version. Again, we averaged the items per first-order dimension and then fit the 10 scale scores to a three-factor model (10:3). The factor loadings from these analyses are summarized in Table 4, and the corresponding model fit indices are summarized in Table 5. The three-factor model yielded an acceptable fit using both the field, $\chi^2(32) = 130.81, p < .01$; CFI = .917, SRMR = .035, PCFI = .652, and laboratory, $\chi^2(32) = 75.24, p < .001$; CFI = .960, SRMR = .033, PCFI = .683, subsamples, along with good parsimony indices. Here again, the three-factor models fit significantly better than single-factor models in both cases: field, $\Delta\chi^2(3) = 123.26$, $p < .001$; $\chi^2(35) = 254.07$, $p < .001$; CFI = .816, SRMR = .066, PCFI = .635; laboratory, $\Delta\chi^2(3) = 56.81, p < .001$; $\chi^2(35) = 132.05, p < .001$; CFI = .910, SRMR = .044, PCFI = .708. All first-order subscales loaded significantly ($p < .001$) and >.75 on their intended second-order dimensions. The correlations among the transition, action, and interpersonal dimensions were significant and high (*rs* >.75, $p < .001$) in both of these analyses. Again, these findings are consistent with both Marks et al.'s (2001) higher-order framework and LePine et al.'s (2008) meta-analytic–based CFA analyses.

Combining the 30-item subsamples ($N = 207$ teams) provided sufficient degrees of freedom to test the higher-order factor structure at the item level. Fitting the 30 items to a 10-factor first-order model (30:10) yielded a deficient overall fit, $\chi^2(360) = 1{,}238.33$, $p < .001$; CFI = .883, SRMR = .061, PCFI = .731, on the basis of the CFI but not SRMR. Notably, the item loadings in this analysis were all >.73, $p < .001$, on their intended constructs. The 30:10 model fit was significantly better than single-factor (30:1) model, which illustrated a poor fit, $\Delta\chi^2(45) = 1{,}154.66$, $p < .001$; $\chi^2(405) = 2{,}392.99$, $p < .001$; CFI = .738, SRMR = .072, PCFI = .685. Fitting the 10 first-order constructs to the three second-order dimensions (30:10:3) yielded a significantly worse model fit, $\Delta\chi^2(32) = 134.54$, $p < .001$; $\chi^2(392) = 1{,}372.87$, $p < .001$; CFI = .870, SRMR = .065, PCFI = .786, although the first-order dimensions all loaded significantly and >.83, $p < .001$, on their intended second-order constructs. Here again, although the CFI is below conventional levels of acceptability, the SRMR value suggests an acceptable fit, and the parsimony index was higher. As was seen with the 50-item version, the 30:10:3 second-order structure fit significantly better fit, $\Delta\chi^2(3) = 209.99$, $p < .001$, than a 30:10:1 structure, $\chi^2(395) = 1{,}582.86$, $p < .001$; CFI = .842, SRMR = .076, PCFI = .765. The second-order variables were highly correlated (mean $r = .88$, $p < .001$; see Table 6), so we fit a 30:10:3:1 model, the results of which are shown in Figure 2. Consistent with our 50:10:3:1 findings and LePine et al.'s (2008) meta-analysis, the first- and second-order loadings were all significant (>.80, $p < .001$), as were the second- and third-order loadings (>.85, $p < .001$).

Of course, the 30-item version of the scale is included in the 50-item version. Accordingly, we used all 714 teams to reassess the fit of Marks et al.'s (2001) higher-order model using the 30 items. The 30:10 first-order model evidenced an acceptable fit, $\chi^2(360) = 1{,}917.43$, $p < .001$; CFI = .925, SRMR = .051, PCFI = .766, and was significantly better than a 30:1 model, $\Delta\chi^2(45) = 6{,}116.20$, $p < .001$; $\chi^2(405) = 8{,}033.63$, $p < .001$; CFI = .634, SRMR = .107, PCFI = .590. The item loadings in this analysis were all >.70, $p < .001$, on their intended constructs. Fitting the 30:10:3 higher-order model evidenced a significantly worse fit, $\Delta\chi^2(32) = 510.41$, $p < .001$, although the 30:10:3 model did evidence acceptable fit indices and an excellent parsimony index, $\chi^2(392) = 2{,}427.84$, $p < .001$; CFI = .905, SRMR = .061, PCFI = .816, with all first-order dimensions loading significantly and >.80, $p < .001$, on their intended second-order constructs in this model. The 30:10:3 second-order structure fit significantly better fit, $\Delta\chi^2(3) = 775.33$, $p < .001$, than the 30:10:1 structure, $\chi^2(395) = 3{,}203.17$, $p < .001$; CFI = .865, SRMR = .093, PCFI = .785.

Summary of 30-item measure. In sum, although the coverage of the 30-item version is less than that of the 50-item version, the shortened subscales exhibited good psychometric properties, and our results are consistent with Marks et al.'s (2001) higher-order framework. Interestingly, using the unified sample, the 30-item version of the measure evidenced an acceptable fit using both the 10 subscales (i.e., 10:3 model) and 30 items (i.e., 30:10:3) as indicators. Whereas the item-level 30:10 model evidenced significantly better fit, the corresponding 30:10:3 model exhibited superior parsimony. Marks et al. anticipated that might be the case, and they suggested that whether researchers should focus on the more specific first-order (30:10) or second-order (30:10:3) dimensions would depend on the purpose of their investigation.

10-Item Version

There may well be some instances when scholars or practitioners may wish to administer a very short measure as a check of current team functioning (Welbourne, 2016). For instance, even the 30-item version would be taxing to answer in a daily experience sampling study design. Accordingly, we used our judgment and input from subject matter experts in two applications and identified what we believed was the most representative single item for each of the 10 first-order dimensions (designated with an * in Appendix A. Using the $N = 714$ unified sample, the psychometric properties of these items at the three second-order dimensions were: transition median $r_{wg} = .87$, ICC1 = .27, ICC2 = .58, $\alpha = .83$; action median $r_{wg} = .89$, ICC1 = .26, ICC2 = .54, $\alpha = .82$; and interpersonal median $r_{wg} = .86$, ICC1 = .29, ICC2 = .58, $\alpha = .85$.

We then fit a three-factor CFA model to the 10 items, which evidenced an acceptable fit, $\chi^2(32) = 286.70$, $p < .001$; CFI = .937, SRMR = .045, PCFI = .666, which was significantly better, $\Delta\chi^2(3) = 462.47$, $p < .001$, than a deficient single-factor CFA, $\chi^2(35) = 749.16$, $p < .001$; CFI = .823, SRMR = .081, PCFI = .640. All items loaded on their respective latent constructs significantly and >.59, $p < .001$.

Discriminant Validity

Team Processes Versus Emergent States

The analyses chronicled previously suggest that the 50-, 30-, and 10-item versions of the measure appeared to have acceptable psychometric properties and are consistent with Marks et al.'s (2001) team process framework. Recall that Marks et al. were interested in the dynamic nature of team members' behaviors—namely, interactions between them and the task. To maintain that focus, they differentiated processes (i.e., action verbs) from other types of mediators in team effectiveness models, such as affective and cognitive emergent states (i.e., adjectives). They argued that the latter

> types of constructs do not denote interaction processes but, instead, tap qualities of a team that represent member attitudes, values, cognitions, and motivations. We prefer to label these types of variables as "emergent states": constructs that characterize properties of the team that are typically dynamic in nature and vary as a function of team context, inputs, processes, and outcomes. Emergent states describe cognitive, motivational, and affective states of teams, as opposed to the nature of their member interaction. (Marks et al., 2001, p. 357)

Emergent states represent the "state" of a team at a particular moment and can be conceived of as antecedents, correlates, or consequences of processes—depending on the design, measurement, and timing of an investigation. Many eloquent models have been advanced that detail the co-evolution of teams' processes and emergent states (e.g., Kozlowski & Ilgen, 2006; Waller, Okhuysen, & Saghafian, 2016). The core theme is that team members engage in different behaviors during any given period of time (e.g., during transition or action phases), which give rise to members' cognitions (e.g., shared mental models, loss of situational awareness) and feelings (e.g., cohesion, affective tone). Processes and states are therefore closely related and co-evolve over time. For instance, engaging in mission analysis and strategy formulation (processes) should give rise to members' shared mental models and strategic intent (states). Monitoring systems and progress toward goal achievement (processes) should yield situational awareness (states). Engaging in different forms of motivation or conflict management (processes) should impact members' collective efficacy and experienced conflict (states). Whereas we encourage future research that conceptualizes and considers the co-evolution of team processes and states, we need validated measurement tools that can index the different types of constructs and model their relationships over time.

Whereas it may be conceptually appealing to differentiate team members' behaviors from their cognitions and feelings, in practice, it may be difficult to distinguish the different types of constructs. This is particularly likely to be the case when team processes and states are measured at the same time using similar methods. LePine et al.'s (2008) meta-analyses estimated that the population correlations between the three second-order team processes and (a) team cohesion ranged from $r = .53$ to .61, $p < .001$, and (b) team potency ranged from $r = .63$ to .70, $p < .001$. These correlations are high enough to warrant the examination of the discriminant validity of the measures we are offering with related team states. For instance, Mathieu et al. (2006) obtained a correlation of $r = .85$, $p < .001$, between estimated latent constructs of overall team processes and team empowerment, and Rapp et al. (2016) found an $r = .89$, $p < .001$, between the two latent constructs. Naturally, the more conceptually similar and highly correlated two latent variables are, the more challenging it is to demonstrate discriminant validity of measures of them. Moreover, team empowerment and processes offer parallel comparisons given that they are both second-order constructs with four and three lower-order indicators, respectively. This makes team empowerment a valuable construct to contrast against team processes. Accordingly, we provide a test of the discriminant validity of our measures of team processes and team empowerment. As team empowerment (Kirkman & Rosen, 1999) is a well-established measure of a popular emergent state, it is a valuable basis for comparison.

Building on the work of Spreitzer (1995), Kirkman and Rosen (1999) advanced the concept of team empowerment as consisting of four facets: (a) *potency*—a collective belief by team members that they can be effective; (b) *meaningfulness*—the tasks that the team works on are important, valuable, and worthwhile; (c) *autonomy*—the team has discretion over their work; and (d) *impact*—the work performed by the team is significant and advances organizational objectives. Kirkman and Rosen developed a 12-item measure of it that has proven to be reliable and valid. Kirkman, Rosen, Tesluk, and Gibson (2004); Maynard, Mathieu, Gilson, O'Boyle, and Cigularov (2013); and others have positioned this conception of team empowerment as an emergent state as defined by Marks et al. (2001). Accordingly, in the following, we report on the discriminant validity of our 30-item measure of team process and Kirkman and Rosen's 12-item measure of team empowerment using a sample of global virtual teams. We then report on the discriminant validity of our 10-item measure of team processes and Kirkman and Rosen's 12-item empowerment measure using a sample of healthcare employees.

Global Virtual Teams and 30-Item Version

Sample

Organizational communities of practice members (which operated as global virtual teams) from a Fortune 100 U.S.-based multinational mining and minerals processing firm with over 300 operations in 44 countries were sampled as part of a larger investigation (Kirkman, Mathieu, Cordery, Kukenberger, & Rosen, 2011). For current purposes, we use team members' survey responses to our 30-item measure of team processes and Kirkman and Rosen's (1999) 12-item measure of team empowerment. In total, useable online surveys were available for 459 respondents from 64 teams. The sample was 78% men with an average age of 41 years ($SD = 9.6$).

Measures

Team processes. Given the sample size, we indexed team processes by averaging the respective 30 item responses according to the three second-order constructs and obtained acceptable psychometric properties parallel to what we reported previously: transition median $r_{wg} = .91$, ICC1 = .09, ICC2 = .38, $\alpha = .96$; action median $r_{wg} = .92$, ICC1 = .08, ICC2 = .35, $\alpha = .91$; and interpersonal median $r_{wg} = .91$, ICC1 = .10, ICC2 = .417, $\alpha = .95$.

Team empowerment. Kirkman and Rosen's (1999) measure includes three items each for the four empowerment dimensions, including *potency* (e.g., "My Community can get a lot done when it works hard"), *meaningfulness* (e.g., "My Community believes that its projects are significant"),

Table 7. Variables Descriptive Statistics and Correlations for Discriminant Validity Samples.

	1	2	3	4	5	6	7	Global Virtual Teams Mean	*SD*	Health Care Teams Mean	*SD*
1. Transition process	1.00	.83	.65	.75	.70	.74	.74	2.96	0.44	4.36	0.33
2. Action process	.97	1.00	.62	.78	.64	.66	.64	2.83	0.44	4.33	0.36
3. Interpersonal process	.85	.88	1.00	.54	.73	.70	.73	3.10	0.47	3.58	0.50
4. Empowerment-potency	.64	.70	.74	1.00	.64	.65	.61	5.43	0.62	4.46	0.32
5. Empowerment-meaningfulness	.60	.57	.64	.70	1.00	.88	.88	5.69	0.52	3.96	0.49
6. Empowerment-autonomy	.52	.59	.60	.61	.46	1.00	.87	4.86	0.56	4.03	0.42
7. Empowerment-impact	.59	.62	.61	.73	.85	.60	1.00	5.63	0.52	3.86	0.53

Note: All correlations are $p < .01$. Values above diagonal are from the global virtual teams sample ($N = 64$), and the ones below are from the health care teams sample ($N = 210$).

Table 8. Discriminant Validity of Team Process and Empowerment Measures.

	Global Virtual Teams ($N = 64$)		Health Care Teams ($N = 210$)	
Subscale Indicators	Process	Empowerment	Process	Empowerment
Transition	.94		.94	
Action	.97		.93	
Interpersonal	.91		.93	
Potency		.82		.92
Meaningfulness		.88		.89
Autonomy		.65		.72
Impact		.93		.83

Note: Table values are standardized confirmatory factor analysis loadings, and all $p < .001$.

autonomy (e.g., "My Community makes its own choices without being told by management"), and *impact* (e.g., "My Community performs tasks that matter to this company"). Psychometric properties for the subscales were: potency median $r_{wg} = .86$, ICC1 = .07, ICC2 = .30, $\alpha = .91$; meaningfulness median $r_{wg} = .89$, ICC1 = .03, ICC2 = .17, $\alpha = .93$; autonomy median $r_{wg} = .80$, ICC1 = .03, ICC2 = .15, $\alpha = .78$; and impact median $r_{wg} = .87$, ICC1 = .04, ICC2 = .22, $\alpha = .88$.

Results

Given the limited sample size, we used scale scores of the three second-order team processes as indicators of an overall team process latent construct and the four empowerment subscales as indicators of an overall empowerment latent construct. A correlation matrix and descriptive statistics for these seven indicators are shown in Table 7.

A two-factor CFA yielded an acceptable model fit, $\chi^2(13) = 43.96$, $p < .001$; CFI = .930, SRMR = .062, PCFI = .576, that was significantly better, $\Delta\chi^2(1) = 59.77$, $p < .001$, than a deficient single factor model, $\chi^2(14) = 103.73$, $p < .001$; CFI = .800, SRMR = .090, PCFI = .533. Table 8 presents the standardized factor loadings that were all significant ($p < .001$) and loaded $\geq .65$ on their respective latent constructs. The correlation between the two latent constructs was $r = .74$, $p < .001$. Although highly correlated, these results support the discriminant validity of the measures of the two constructs.

Health Care Teams and 10-Item Version

Sample

Health care employees from five different facilities in the mid-Atlantic region of the United States were sampled as part of a larger investigation (D'Innocenzo, Luciano, Mathieu, Maynard, & Chen, 2016). For current purposes, we use their survey responses to our 10-item measure of team processes and Kirkman and Rosen's (1999) 12-item measure of team empowerment. In total, useable surveys were available for 1,449 respondents who were members of 210 teams. The sample was 83% women and had an average age of 44.7 years ($SD = 11.4$).

Measures

Team processes. As previously described, the 10 items were used as indicators of the three second-order team process constructs and evidenced acceptable psychometric properties. Specifically, transition median $r_{wg} = .89$, ICC1 = .19, ICC2 = .56, $\alpha = .95$; action median $r_{wg} = .91$, ICC1 = .18, ICC2 = .55, $\alpha = .87$; and interpersonal median $r_{wg} = .87$, ICC1 = .21, ICC2 = .60, $\alpha = .93$.

Team empowerment. We again used Kirkman and Rosen's (1999) 12-item measure of empowerment. Psychometric properties for the subscales were: potency median $r_{wg} = .93$, ICC1 = .11, ICC2 = .42, $\alpha = .86$; meaningfulness median $r_{wg} = .92$, ICC1 = .10, ICC2 = .46, $\alpha = .93$; autonomy median $r_{wg} = .79$, ICC1 = .17, ICC2 = .53, $\alpha = .87$; and impact median $r_{wg} = .94$, ICC1 = .05, ICC2 = .24, $\alpha = .92$.

Results

We again used scale scores of the three second-order team processes as indicators of an overall team process latent construct and the four empowerment subscales as indicators of an overall empowerment latent construct. A correlation matrix and descriptive statistics for these seven indicators are also shown in Table 7. The two-factor CFA revealed an excellent model fit, $\chi^2(13) = 75.72$, $p < .001$; CFI = .958, SRMR = .051, PCFI = .593, that was significantly better, $\Delta\chi^2(1) = 130.23$, $p < .001$, than a single-factor model, $\chi^2(14) = 205.95$, $p < .001$; CFI = .872, SRMR = .065, PCFI = .581. The standardized factor loadings were all significant ($p < .001$) on their respective latent constructs and >.80 (see Table 8). The correlation between the two latent constructs was $r = .83$, $p < .001$.

Discussion

Our overall goal for this work was to develop and provide construct validity evidence for a measure of team processes aligned with Marks et al.'s (2001) theoretical framework. After generating 50 items, 5 for each of Marks et al.'s 10 first-order dimensions, we gathered content validity judgments from SMEs. They categorized the items according to their intended dimensions in all but a few cases. Then, using data from both field and laboratory teams, we conducted item analyses and evaluated member agreement as the basis for aggregating their responses to represent team-level indices. The data supported aggregation, and the remaining analyses were conducted using average member responses per team (see Chen et al., 2004).

Next, we tested the unidimensionality of the five-item scales, and they evidenced well-fitting single-factor CFAs in all instances. All items exhibited significant and large loadings in these analyses, including the items that were questionable from the content validity categorizations. Therefore, we created substantive-based five-item parcels for the first-order dimensions and tested

Marks et al.'s (2001) higher-order factor structure. The 10 first-order dimensions well fit to their hypothesized three second-order (10:3) transition, action, and interpersonal dimensions using both the field and laboratory subsamples. Combining the two samples, we then fit the 50 items to a 10-dimensional first-order model (50:10) as well as a second-order (50:10:3) CFA model whereby we mapped the 10 first-order dimensions to their corresponding three second-order dimensions. Both the 10-dimensional first-order model (50:10) and the second-order model (50:30:10) evidenced mixed results, surpassing conventional fit indices thresholds for the SRMR but not the CFI. All item and dimensional loadings were significant in all instances. And although, as one would expect, the 50:10:3 second-order model fit significantly worse than the 50:10-dimensional first-order model, the former fit significantly better than a 50:10:1 second-order model. The higher-order 50:30:3 model also exhibited a significantly higher parsimony index than the 50:10 model. All totaled, there appeared to be support for the use of the 50-item measure and evidence consistent with Marks et al.'s higher-order framework.

Although the 50-item version of the measure evidenced high content validity and anticipated factor structures, it was not completely supported, and many researchers would prefer to use shorter forms. Accordingly, we reproduced our analyses using additional field and laboratory samples who completed a 30-item shorter version (3 items per first-order dimension). The findings for the 30-item version paralleled those obtained for the longer version. Moreover, we provided evidence that a 10-item version (1 item per first-order dimension) could fit Marks et al.'s (2001) three-dimensional higher-order framework. Last, we provided evidence of the discriminant validity of the 30-item and 10-item versions of our team process measures versus a popular measure of team empowerment (Kirkman & Rosen, 1999) using a sample of global virtual knowledge teams and a sample of health care teams, respectively. Whereas our analyses suggested that our measures of team processes are distinguishable from Kirkman and Rosen's (1999) measures of empowerment, the two latent constructs were still highly correlated. Future research should be devoted to testing the unique and combined relationships of both constructs as related to important criteria such as team performance and viability.

Which Version Should be Used?

Given the findings chronicled previously, the natural inclination might be to simply use the 10-item version of the measure. However, that decision is not likely appropriate in all circumstances. The longer versions of the scales offer a more thorough representation of the construct domains. Marks et al. (2001) limited their taxonomy to 10 dimensions, which necessarily meant that they subsumed several subthemes within their first-order dimensions. For example, they suggested that strategy formulation and planning included not only deliberate planning processes but also the specification of contingency plans—both most typically done during transition periods. They further suggested that reactive strategy adjustment, or improvisation, would also be included in this dimension and would more likely be exhibited during action phases. In other words, even the 10 first-order dimensions are multifaceted constructs that would not likely be thoroughly represented using a single item or two.

We believe that decisions about which version to use should be informed by the larger purpose of an investigation. For instance, Marks et al. (2001) suggested that measuring all 10 first-order dimensions, or perhaps direct measures of the three second-order dimensions, would be advisable when the goal is to gather a comprehensive view of team processes. The 50- or 30-item versions would be suitable in such instances. Alternatively, if the research goal is targeted at a particular first-order dimension, then perhaps using the 5-item scale for that specific process would be advisable. Notably, Marks et al.'s recommendations were elaborated by LePine et al. (2008), who advised:

> When theory focuses on relationships with the overall quality of team processes, hypotheses could be tested with broad measures of teamwork. To avoid deficiency, such measures should either consist of items written to tap the general concept itself or a representative array of items from the narrower teamwork processes. When theory focuses on relationships with a more specific aspect of teamwork, or teamwork in the context of a specific phase of a team's existence, hypotheses could be tested using appropriate midlevel measures of teamwork (transition, action, interpersonal processes). Such measures could include items written to tap the appropriate first-order concept directly or include items from narrower measures that correspond to the first-order measure. Finally, if the theory is focused on understanding fine-grained interpersonal activities and drawing distinctions among them, then hypotheses should be tested with measures that tap the narrowest teamwork processes. (p. 294)

We concur with the aforementioned recommendations when it comes to research purposes. Yet applied considerations may impose other decision criteria. For instance, "pulse surveys" are gaining in popularity and are designed to provide quick insights as to the current functioning of a team (Welbourne, 2016). These measures are gathered at a high frequency—perhaps weekly or even daily—and are intended to serve as leading indicators of team status. Clearly 50-item and even 30-item surveys would prove too taxing and lead to survey fatigue if used in this fashion, but a 10-item version may be suitable. Whereas the 10 items that we selected for such use exhibited acceptable second-order psychometric properties, other subsets may be more suitable in other circumstances. In other words, initial grounding efforts may suggest other items are better aligned with organizational issues or pressure points. Given the unidimensionality of the 5-item sets, we would anticipate that the higher-order structure of different item subsets would likely be supported. It is also the case that additional items may be warranted to more thoroughly sample the construct domain of any of the first- or second-order constructs. Furthermore, the CFI indices suggest that there is room for improvement in the overall models.

In sum, the answer to the "which version" question is to seek the proper measurement fit. Luciano et al. (2018, p. 600) proposed that "measurement fit reflects the degree of alignment between how a construct is *conceptualized* and *measured.*" They further suggested "that achieving measurement fit requires an iterative approach involving three core components: (a) construct elements, (b) measurement features, and (c) contextual considerations." The measures that we offer herein, however, afford the opportunity to choose the version that is best aligned with one's circumstances and purposes.

When Should the Measures Be Collected?

The modal research design employed by the studies summarized in Table 1 was to gather survey measures of all team processes simultaneously. Certainly, this is a relatively more convenient way to gather data, particularly if there are limited opportunities to sample respondents. It also yields a snapshot of activities that is suitable for recurring surveys used for feedback and development purposes. Yet doing so requires respondents to mentally aggregate different experiences over time and draw a general inference as to how well they plan activities, coordinate actions, manage conflict, and so on, over time. However, Marks et al.'s (2001) model was predicated on the notion that teams execute different processes during versus between performance episodes. Gathering data at a single time therefore necessarily means that some constructs are being assessed at times other than when they have occurred. For instance, suppose a project design team is surveyed in the midst of their development of a prototype (i.e., action processes). If their proficiencies at monitoring progress toward goal achievement and coordinating their efforts are wanting, that may well be reflected in members' lower survey scores. But they also may report relatively poor transition processes because either (a) they in fact rushed their mission analysis, goal specification, and strategy development processes or (b) they are engaging in attributional processes given their current action process struggles (cf. Martell, Guzzo, & Willis, 1995; Staw, 1975). It is virtually impossible to disentangle those two explanations, and data would be far more informative if the different processes were measured when the processes were occurring.

We echo Marks et al.'s (2001) recommendation that "measures should be gathered at appropriate times and using measures that are most suitable for the nature of the construct(s) being examined—all based on the knowledge garnered from a time-sensitive team task analysis" (p. 371). LePine et al. (2008) reiterated this logic, submitting "the point of this discussion is that we encourage future research to employ time-based research designs whereby measures of different team processes can be aligned with when they are anticipated to occur" (p. 297). Notably, the two field samples that administered the 50-item measures collected the transition processes scales weeks before they collected the action and interpersonal process scales. The correlations between transition processes and others were markedly lower in those two samples than they were in other samples that collected all measures concurrently.

Fisher (2014) sampled student teams competing in a business simulation over time and aligned the measurement of team processes with when they were presumed to occur. He assessed planning processes early in the simulation and then interpersonal and action processes during later rounds. Given that teams may evolve at different rates, a time-based sampling approach may well entail an ipsative design where different teams are completing measures at different times. For instance, Maynard (2007) examined accounting teams performing audits of varying duration. He monitored the individual teams and collected measures of transition processes as each team shifted from one stage of the audit to another and measures of action and interpersonal processes during each stage. Logistically, this design requires either some sort of tracking mechanism that can monitor when teams are operating in different phases or perhaps an experience sampling style design whereby members are completing short surveys (e.g., a 10-item version) on a frequent basis. An experience sampling style design may be particularly advantageous when teams are performing multiple tasks concurrently that are likely to progress at different rates.

Temporally oriented research designs also raise an important framing issue for measures of team processes—and for that matter, for any variable that varies over time. Researchers need to specify a temporal window that respondents should consider when answering survey questions about their team processes (Luciano et al., 2018). In laboratory investigations that collect measures at different times, it may be clear as to when team members were planning versus executing actions. But in many instances, teams may have performed multiple performance episodes, and survey questions should specify whether respondents should focus on some specific instance or mentally aggregate over time. Even then, including temporal anchors in instructions such as "over the past 3 months . . . " or "since the previous survey administration . . . " will help to focus respondents and minimize halo-type effects.

Temporal designs are also taxing and can easily lead to survey burnout and response biases from repeated administrations. Therefore, researchers might consider sampling different subsets of team members at different times as key informants to reduce survey fatigue and potential testing-related threats to internal validity. This would be a viable option if members are presumed (or previously demonstrated) to have high agreement concerning their team processes. In any case, assessing processes that are presumed to develop over time during a single administration is likely to misalign measures and phenomena, trigger attributional biases, inflate methods-related effects, and undermine measurement alignment (see Luciano et al., 2018).

What Other Techniques Should Be Considered?

Surveying team members is often a convenient way to gather measures of team processes. Members are well informed to provide insights regarding their interpersonal processes. In fact, members are likely one of the best sources for assessing interpersonal constructs. However, other sources of information may be preferable for measuring transition and action processes. Transition processes often yield by-products of members' efforts such as strategic planning documents, charters, and specified goals. Actions processes (e.g., actions among surgical crews) are perhaps better measured using live observers or raters watching video recordings than they could be from members in the middle of working on a patient. For instance, Marks et al. (2005) had teams complete a mission analysis and planning phase after which team leaders were interviewed concerning their transition processes. Teams then "flew" a laboratory flight simulation that was recorded and later coded in terms of action processes by trained observers. Kamphuis et al. (2011) had trained coders review communication logs to index team transition processes and the distribution of information in emails to different team members as a measure of action processes. Notably, the correlations between the different processes in Marks et al. and Kamphuis et al. were substantially lower than those reported in LePine et al.'s (2008) meta-analysis or herein, no doubt in part because they were less susceptible to methods effects.

Survey responses are valuable indices of team processes, but here again, we echo the advice of Marks et al. (2001) and LePine et al. (2008) to consider alternative methods of measurement. Fortunately, recent developments may enable scholars and practitioners to finally act on such advice. For example, Kozlowski (2015) submitted that "researchers should routinely seek to supplement questionnaire-based assessments with alternative measures of behavior" (p. 285) and went on to describe how video-based and behavioral trace measures can be leveraged. More generally, Luciano and colleagues (2018) detailed how different data streams such as verbal communication (i.e., both what is said and speech patterns among members), behaviors (e.g., motion, gestures, posture), and physiological responses (e.g., respiration and heart rates, electrocardiograms, blood pressure) may be used to index team processes and other constructs. In short, newer methods of measurement can complement surveys and provide additional insights into the dynamics of team processes while minimizing certain method effects and intrusiveness. Furthermore, these team process scales may prove useful in validating the measurements constructed from data streams generated by emergent technologies. However, we wish to emphasize that all methods have limitations and encourage researchers and practitioners to engage in a measurement fitting process (cf. Kozlowski, 2015; Luciano et al., 2018).

Conclusion

Marks et al. (2001) advanced a framework of team processes that has proven to be quite popular but not systematically investigated. We submit that is attributable in part to the lack of a validated measure aligned with their theory. Accordingly, we developed and demonstrated content and construct validity evidence of a survey measure of team processes. We illustrated the properties of 50-, 30-, and 10-item versions of the measure. Moreover, using data from over 700 teams, we tested Marks et al.'s higher-order model and found results that were largely consistent with both their 10 first-order dimensions and the 10:3 second-order framework. Additional analyses with samples of global virtual knowledge teams and health care employees provided evidence of the discriminant validity of the team process measures versus a measure of team empowerment. Notably, we intentionally subjected these scales to the various tests to provide future users with information concerning the psychometric properties of different configurations of the measures for different potential uses. We are not suggesting that these varied tests are required to be performed by all future users as few opportunities afford sufficient sample sizes to do so. We do recommend that users report agreement and aggregate reliabilities for the scale configuration that they employ so that additional evidence concerning the construct validity of the scales can be accumulated. Our hope is that these measures provide a useful common metric for team scholars and practitioners going forward, and we encourage their use as part of a larger portfolio of measures of team process.

Appendix A: Team Process Scale Items

Transition Processes

To what extent does our team actively work to . . .

Mission Analysis

1. Identify our main tasks?
*2. Identify the key challenges that we expect to face?
3. Determine the resources that we need to be successful?
4. Develop a shared understanding of our purpose or mission?
5. Understand the needs of our primary stakeholders (e.g., customers, top management, other organizational units)?

Goal Specification

6. Set goals for the team?
*7. Ensure that everyone on our team clearly understands our goals?
8. Link our goals with the strategic direction of the organization?
9. Prioritize our goals?
10. Set specific timelines for each of our goals?

Strategy Formulation and Planning

*11. Develop an overall strategy to guide our team activities?
12. Prepare contingency ("if-then") plans to deal with uncertain situations?
13. Know when to stick with a given working plan, and when to adopt a different one?
14. Periodically re-evaluate the quality of our working plans?
15. Specify the sequence in which work products should be accomplished?

Action Processes

To what extent does our team actively work to . . .

Monitoring Progress Toward Goals

16. Regularly monitor how well we are meeting our team goals?
17. Use clearly defined metrics to assess our progress?
*18. Seek timely feedback from stakeholders (e.g., customers, top management, other organizational units) about how well we are meeting our goals?
19. Know whether we are on pace for meeting our goals?
20. Let team members know when we have accomplished our goals?

Systems Monitoring

21. Monitor and manage our resources (e.g., financial, equipment, etc.)?
*22. Monitor important aspects of our work environment (e.g., inventories, equipment and process operations, information flows)?

(continued)

23. Monitor events and conditions outside the team that influence our operations?
24. Ensure the team has access to the right information to perform well?
25. Manage our personnel resources?

Team Monitoring and Backup

26. Develop standards for acceptable team member performance?
27. Balance the workload among our team members?
*28. Assist each other when help is needed?
29. Inform team members if their work does not meet standards?
30. Seek to understand each other's strengths and weaknesses?

Coordination

31. Communicate well with each other?
32. Smoothly integrate our work efforts?
*33. Coordinate our activities with one another?
34. Re-establish coordination when things go wrong?
35. Have work products ready when others need them?

Interpersonal Processes

To what extent does our team actively work to . . .

Conflict Management

*36. Deal with personal conflicts in fair and equitable ways?
37. Show respect for one another?
38. Maintain group harmony?
39. Work hard to minimize dysfunctional conflict among members?
40. Encourage healthy debate and exchange of ideas?

Motivating and Confidence Building

41. Take pride in our accomplishments?
42. Develop confidence in our team's ability to perform well?
*43. Encourage each other to perform our very best?
44. Stay motivated, even when things are difficult?
45. Reward performance achievement among team members?

Affect Management

46. Share a sense of togetherness and cohesion?
47. Manage stress?
*48. Keep a good emotional balance in the team?
49. Keep each other from getting overly emotional or frustrated?
50. Maintain positive work attitudes?

Note: The first three items listed under each subscale represent the 30-item shorter form.
The * items represent the 10-item short form.
Response scale: 1 = *not at all*; 2 = *very little*; 3 = *to some extent*; 4 = *to a great extent*; 5 = *to a very great extent*.

Authors' Notes

We thank Michelle Marks for her contributions during the item generation phase of this work. We also thank John Cordery, Leslie DeChurch, Elar Killumets, Bradley Kirkman, Glenn Littlepage, Nora Madjar, M. Travis Maynard, Ronald Piccolo, Timothy Vogus, and Jessica Wildones for generously sharing their data for current purposes. Finally, we thank the dozens of anonymous organizational scholars who participated in the item sorting task. Additional information and resources concerning these measures can be found at https://UCHuskypack.com.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was conducted with support from the United States Army Research Institute 624 Organizational Research Methods 21(3) (Contract: W911NF-15-1-0014; The Development of Construct Validation of Unobtrusive Dynamic Measures of Team Process and Emergent States). The opinions expressed herein however, are strictly ours.

ORCID iD

John E. Mathieu https://orcid.org/0000-0002-6881-4022

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Specifically, in a series of model tests, we collapsed the items from the two most highly correlated latent variables in each sample into a single factor and then compared the three-factor to a two-factor model. In all instances, the two factors models exhibited significantly ($p < .01$) worse fits. Details available from the first author.
2. The model fit indices for this model are identical to those of the 50:10:3 model because there are three second-order latent constructs.
3. Details available from the first author.
4. Note that with three indicators, a single-factor confirmatory factor analysis is just identified and therefore does not yield overall model fit indices.

References

Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology*, *76*(5), 732-740.

Bagozzi, R. P., & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods*, *1*(1), 45-87.

Baumann, M. R. (2013). Where can I publish "Group" research? *INGroup News Letter*, *3*, 2-8.

Behfar, K. J., Peterson, R. S., Mannix, E. A., & Trochim, W. M. K. (2008). The critical role of conflict resolution in teams: A close look at the links between conflict type, conflict management strategies, and team outcomes. *Journal of Applied Psychology*, *93*(1), 170-188. doi:10.1037/0021-9010.93.1.170

Bliese, P. D., Maltarich, M. A., Hendricks, J. L., Hofmann, D. A., & Adler, A. B. (2018). Improving the measurement of group-level constructs by optimizing between-group differentiation. *Journal of Applied Psychology*, *104*(2), 293-302.

Chen, G., Bliese, P. D., & Mathieu, J. E. 2004. Conceptual framework and statistical procedures for delineating and testing multilevel theories of homology. *Organizational Research Methods*, *8*(4), 375-409.

Chen, G., Kanfer, R., DeShon, R. P., Mathieu, J. E., & Kozlowski, S. W. J. (2009). The motivating potential of teams: Test and extension of Chen and Kanfer's (2006) cross-level model of motivation in teams. *Organizational Behavior and Human Decision Processes*, *110*(1), 45-55. doi:10.1016/j.obhdp.2009.06.006

Cheung, M. W.-L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods*, *10*(1), 40-64.

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, *86*(3), 386-400.

Courtright, S. H., McCormick, B. W., Mistry, S., & Wang, J. (2017). Quality charters or quality members? A control theory perspective on team charters and team performance. *Journal of Applied Psychology*, *107*, 1462-1470. doi:10.1037/apl0000229

De Jong, B. A., & Elfring, T. (2010). How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring, and effort. *Academy of Management Journal*, *53*(3), 535-549. doi:10.5465/amj.2010.51468649

DeChurch, L. A., & Haas, C. D. (2008). Examining team planning through an episodic lens: Effects of deliberate, contingency, and reactive planning on team effectiveness. *Small Group Research*, *39*(5), 542-568. doi:10.1177/1046496408320048

Digman, J. M. (1997). Higher-order factors of the Big Five. *Journal of Personality and Social Psychology*, *73*(6), 1246-1256.

D'Innocenzo, L., Luciano, M., Mathieu, J. E., Maynard, M. T., & Chen, G. (2016). Empowered to perform: A multi-level investigation of the influence of empowerment on performance in hospital units. *Academy of Management Journal*, *59*(4), 1290-1307.

Drach-Zahavy, A., & Somech, A. (2001). Understanding team innovation: The role of team processes and structures. *Group Dynamics: Theory, Research, and Practice*, *5*, 111-123.

Eddy, E. R., Tannenbaum, S. I., & Mathieu, J. E. (2013). Helping teams to help themselves: Comparing two team-led debriefing methods. *Personnel Psychology*, *66*(4), 975-1008. doi:10.1111/peps.12041

Firth, B. M., Hollenbeck, J. R., Miles, J. E., Ilgen, D. R., & Barnes, C. M. (2015). Same page, different books: Extending representational gaps theory to enhance performance in Multiteam systems. *Academy of Management Journal*, *58*(3), 813-835. doi:10.5465/amj.2013.0216

Fisher, D. M. (2014). Distinguishing between taskwork and teamwork planning in teams: Relations with coordination and interpersonal processes. *Journal of Applied Psychology*, *99*(3), 423-436. doi:10.1037/a0034625

Fisher, D. M., Bell, S. T., Dierdorff, E. C., & Belohlav, J. A. (2012). Facet personality and surface-level diversity as team mental model antecedents: Implications for implicit coordination. *Journal of Applied Psychology*, *97*(4), 825-841. doi:10.1037/a0027851

Gillet, B., & Schwab, D. P. (1975). Convergent and discriminant validities of corresponding Job Descriptive Index and Minnesota Satisfaction Questionnaire scales. *Journal of Applied Psychology*, *60*(3), 313-317.

Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. *Advances in Experimental Social Psychology*, *8*, 45-99.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis: Pearson New International Edition*. Upper Saddle River, NJ: Pearson.

Haynes, S. N., Richard, D. R., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238-247.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55.

Hyatt, D. E., & Ruddy, T. M. (1997). An examination of the relationship between work group characteristics and performance: Once more into the breech. *Personnel Psychology*, *50*(3), 553-585.

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology*, *67*(2), 219-229.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, *69*(1), 85-98.

Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2003). The core self-evaluations scale: Development of a measure. *Personnel Psychology*, *56*(2), 303-331.

Kamphuis, W., Gaillard, A. W. K., & Vogelaar, A. L. W. (2011). The effects of physical threat on team processes during complex task performance. *Small Group Research*, *42*(6), 700-729. doi:10.1177/1046496411407522

Killumets, E., D'Innocenzo, L., Maynard, M. T., & Mathieu, J. E. (2015). A multilevel examination of the impact of team interpersonal processes. *Small Group Research*, *46*(2), 227-259. doi:10.1177/1046496415573631

Kirkman, B. L., Mathieu, J. E., Cordery, J. L., Kukenberger, M. R., & Rosen, B. (2011). Managing a new collaborative entity in business organizations: Understanding organizational communities of practice effectiveness. *Journal of Applied Psychology*, *96*(6), 1234-1245.

Kirkman, B. L., & Rosen, B. (1999). Beyond self-management: Antecedents and consequences of team empowerment. *Academy of Management Journal*, *42*(1), 58-74.

Kirkman, B. L., Rosen, B., Tesluk, P. E., & Gibson, C. B. (2004). The impact of team empowerment on virtual team performance: The moderating role of face-to-face interaction. *Academy of Management Journal*, *47*(2), 175-192.

Kozlowski, S. W. 2015. Advancing research on team process dynamics: Theoretical, methodological, and measurement considerations. *Organizational Psychology Review*, *5*(4), 270-299.

Kozlowski, S. W., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest*, *7*(3), 77-124.

Landis, R. S. (2013). Successfully combining meta-analysis and structural equation modeling: Recommendations and strategies. *Journal of Business and Psychology*, *28*(3), 251-261.

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815-852.

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, *61*(2), 273-307.

Li, A. N., & Liao, H. (2014). How do leader-member exchange quality and differentiation affect performance in teams? An integrated multilevel dual process model. *Journal of Applied Psychology*, *99*(5), 847-866. doi:10.1037/a0037233

Li, N., Zhao, H. H., Walter, S. L., Zhang, X. A., & Yu, J. (2015). Achieving more with less: Extra milers' behavioral influences in teams. *Journal of Applied Psychology*, *100*(4), 1025-1039. doi:10.1037/apl0000010

Lorinkova, N. M., Pearsall, M. J., & Sims, H. P., Jr. (2013). Examining the differential longitudinal performance of directive versus empowering leadership in teams. *Academy of Management Journal*, *56*(2), 573-596. doi:10.5465/amj.2011.0132

Luciano, M. M., Mathieu, J. E., Park, S., & Tannenbaum, S. I. (2018). A fitting approach to construct and measurement alignment: The role of big data in advancing dynamic theories. *Organizational Research Methods*, *21*(3), 592-632.

Marks, M. A., Mathieu, J. E., DeChurch, L. A., Panzer, F. J., & Alonso, A. (2005). Teamwork in multiteam systems. *Journal of Applied Psychology*, *90*(5), 964-971. doi:10.1037/0021-9010.90.5.964

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, *26*(3), 356-376.

Martell, R. F., Guzzo, R. A., & Willis, C. E. (1995). A methodological and substantive note on the performance-cue effect in ratings of work-group behavior. *Journal of Applied Psychology*, *80*(1), 191-195.

Maruping, L. M., Venkatesh, V., Thatcher, S. M. B., & Patel, P. C. (2015). Folding under pressure or rising to the occasion? Perceived time pressure and the moderating role of team temporal leadership. *Academy of Management Journal*, *58*(5), 1313-1333. doi:10.5465/amj.2012.0468

Mathieu, J. E., Gilson, L. L., & Ruddy, T. M. (2006). Empowerment and team effectiveness: An empirical test of an integrated model. *Journal of Applied Psychology*, *91*(1), 97-108. doi:10.1037/0021-9010.91.1.97

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Cannon-Bowers, J. A., & Salas, E. (2005). Scaling the quality of teammates' mental models: Equifinality and normative comparisons. *Journal of Organizational Behavior*, *26*(1), 37-56. doi:10.1002/job.296

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85*(2), 273-283.

Mathieu, J. E., Hollenbeck, J. R., van Knippenberg, D., & Ilgen, D. R. (2017). A century of work teams in the *Journal of Applied Psychology*. *Journal of Applied Psychology*, *102*(3), 452-467.

Mathieu, J. E., Maynard, T. M., Taylor, S. R., Gilson, L. L., & Ruddy, T. M. (2007). An examination of the effects of organizational district and team contexts on team processes and performance: A meso-mediational model. *Journal of Organizational Behavior*, *28*(7), 891-910. doi:10.1002/job.480

Mathieu, J. E., & Schulze, W. (2006). The influence of team knowledge and formal plans on episodic team process-performance relationships. *Academy of Management Journal*, *49*(3), 605-619. doi:10.5465/AMJ.2006.21794678

Mathieu, J. E., & Taylor, S. R. (2006). Clarifying conditions and decision points for mediational type inferences in organizational behavior. *Journal of Organizational Behavior*, *27*(8), 1031-1056.

Mathieu, J. E., & Taylor, S. R. (2007). A framework for testing meso-mediational relationships in organizational behavior. *Journal of Organizational Behavior*, *28*(2), 141-172.

Mathieu, J. E., Wolfson, M. A., & Park, S. (2018). The evolution of work team research since Hawthorne. *American Psychologist*, *73*(4), 308-321.

Maynard, M. T. (2007). *The impact of experience and familiarity: An examination in project teams*. Unpublished doctoral disseration, University of Connecticut, Storrs, CT.

Maynard, M. T., Mathieu, J. E., Gilson, L. L., O'Boyle, E. H. Jr., & Cigularov, K. P. (2013). Drivers and outcomes of team psychological empowerment: A meta-analytic review and model test. *Organizational Psychology Review*, *3*(2), 101-137.

McGrath, J. E. (1964). *Social psychology: A brief introduction*. New York, NY: Holt, Rinehart and Winston.

McKenzie, J. F., Wood, M. L., Kotecki, J. E., Clark, J. K., & Brey, R. A. (1999). Establishing content validity: Using qualitative and quantitative steps. *American Journal of Health Behavior*. *23*, 311-318.

Mulaik, S. A. (1998). Parsimony and model evaluation. *The Journal of Experimental Education*, *66*(3), 266-273.

Pearsall, M. J., & Venkataramani, V. 2015. Overcoming asymmetric goals in teams: The interactive roles of team learning orientation and team identification. *Journal of Applied Psychology*, *100*(3), 735-748.

Rapp, T. L., Bachrach, D. G., Rapp, A. A., & Mullins, R. (2014). The role of team goal monitoring in the curvilinear relationship between team efficacy and team performance. *Journal of Applied Psychology*, *99*(5), 976-987. doi:10.1037/a0036978

Rapp, T. L., Gilson, L. L., Mathieu, J. E., & Ruddy, T. (2016). Leading empowered teams: An examination of the role of external team leaders and team coaches. *Leadership Quarterly*, *27*(1), 109-123. doi:10.1016/j.leaqua.2015.08.005

Rapp, T. L., & Mathieu, J. E. (2007). Evaluating an individually self-administered generic teamwork skills training program across time and levels. *Small Group Research*, *38*(4), 532-555. doi:10.1177/1046496407300479

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, *27*(2), 94-104.

Saavedra, R., Earley, P. C., & Van Dyne, L. (1993). Complex interdependence in task-performing groups. *Journal of Applied Psychology*, *78*(1), 61-72.

Salas, E., Stagl, K. C., Burke, C. S., & Goodwin, G. F. (2007). *Fostering team effectiveness in organizations: Toward an integrative theoretical framework*. Paper presented at the Nebraska Symposium on Motivation.

Sonnentag, S., & Volmer, J. (2009). Individual-level predictors of task-related teamwork processes: The role of expertise and self-efficacy in team meetings. *Group and Organization Management*, *34*(1), 37-66. doi:10.1177/1059601108329377

Spreitzer, G. M. (1995). Psychological empowerment in the workplace: Dimensions, measurement, and validation. *Academy of Management Journal*, *38*(5), 1442-1465.

Staw, B. M. (1975). Attribution of the "causes" of performance: A general alternative interpretation of cross-sectional research on organizations. *Organizational Behavior and Human Performance*, *13*(3), 414-432.

Sui, Y., Wang, H., Kirkman, B. L., & Li, N. (2016). Understanding the curvilinear relationships between LMX differentiation and team coordination and performance. *Personnel Psychology*, *69*(3), 559-597. doi:10.1111/peps.12115

Summers, J. K., Humphrey, S. E., & Ferris, G. R. (2012). Team member change, flux in coordination, and performance: Effects of strategic core roles, information transfer, and cognitive ability. *Academy of Management Journal*, *55*(2), 314-338. doi:10.5465/amj.2010.0175

Sundstrom, E., De Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist*, *45*(2), 120-133.

Tekleab, A. G., Quigley, N. R., & Tesluk, P. E. (2009). A longitudinal study of team conflict, conflict management, cohesion, and team effectiveness. *Group and Organization Management*, *34*(2), 170-205. doi:10.1177/1059601108331218

Waller, M. J., Okhuysen, G. A., & Saghafian, M. (2016). Conceptualizing emergent states: A strategy to advance the study of group dynamics. *The Academy Of Management Annals*, *10*(1), 561-598.

Welbourne, T. M. (2016). The potential of pulse surveys: Transforming surveys into leadership tools. *Employment Relations Today*, *43*(1), 33-39.

Author Biographies

John E. Mathieu is a Board of Trustees professor and the Friar Chair in Leadership and Teams at the University of Connecticut. His interests include models of team and multiteam effectiveness and cross-level models of organizational behavior. He has worked with several Fortune 500 companies, the armed services, federal, state, and public organizations.

Margaret M. Luciano is an assistant professor of management at the W. P. Carey School of Business, Arizona State University. Her research interests include intergroup dynamics, teams, and systems, with a particular interest in health care settings. Her research has been published in several leading peer-reviewed academic journals.

Lauren D'Innocenzo is an assistant professor of organizational behavior at Drexel University's LeBow College of Business. She received her PhD in management from the University of Connecticut. In her research, she focuses on team processes and emergent states, shared leadership, and cross-level models of team effectiveness.

Elizabeth A. Klock is a PhD student in management at the University of Connecticut. She graduated from Washington & Jefferson College with a BA in psychology. Her research interests include team composition, team emergent states and processes, and personality, with a specific focus in understanding optimal configurations of personalities.

Jeffery LePine is a professor and the Petsmart Chair in Leadership at the W.P. Carey School of Business, Arizona State University. His research focuses on team functioning and effectiveness, proactive behavior and adaptability, leadership, engagement, and stress.